

Et si les réseaux sociaux pouvaient nous aider dans nos choix de carrière?

Rémy Kessler**, Guy Lapalme*, Fabrizio Gotti*, Abdessamad Outerqiss*, Philippe Langlais*

*Université de Montréal

C.P. 6128, succursale Centre-ville, Montréal H3C 3J7, Canada
{lapalme,gottif,outerqia,felipe}@iro.umontreal.ca

** IRISA - UMR 6074, Université de Bretagne-Sud
56017 Vannes, France
remy.kessler@univ-ubs.fr

Résumé. Dans cet article, nous présentons une méthode d'analyse de corpus afin de générer deux interfaces originales de visualisation dans le domaine de l'e-recrutement. Notre approche s'appuie sur des millions de profils issus de plusieurs réseaux sociaux et sur des milliers d'offres d'emploi collectées sur Internet. Nous décrivons dans ces travaux les étapes nécessaires pour leur réalisation. La première visualisation est une carte dynamique indiquant les métiers qui recrutent, dans quel domaine, dans quelle région tandis que la seconde met en avant les parcours professionnels et permet d'observer les perspectives ainsi que les antécédents à plus ou moins long terme pour chaque métier considéré.

1 Introduction

Les développements rapides des réseaux sociaux durant la dernière décennie ont considérablement modifié la dynamique de recherche d'emploi, comme le décrivent Sivabalan et al. (2014). Dans le cadre du projet de recherche Butterfly Predictive Project¹ (BPP), nous faisons l'hypothèse que l'acquisition et l'exploitation des traces laissées par les individus sur les réseaux sociaux (LinkedIn, Viadeo, etc.) sont une voie intéressante afin de construire des ressources permettant d'aider au positionnement professionnel des candidats et ainsi faciliter leur mise en correspondance avec des emplois à pourvoir. Chaque jour, des milliers d'offres d'emploi sont diffusées sur Internet en même temps que des milliers de profils candidats sont créés ou modifiés sur les sites d'emploi ou les réseaux sociaux, aussitôt indexés par des métamoteurs spécialisés ou des moteurs de recherche généralistes. Plusieurs organismes privés ou publics publient régulièrement des statistiques permettant de mesurer les tendances. On citera par exemple Keljob², l'United States Department of Labor³ ou Statistique Canada⁴. Il ne s'agit cependant que d'*instantanés* reflétant l'état du marché. Nous présentons dans cet article

1. <http://rali.iro.umontreal.ca/rali/?q=fr/butterfly-predictive-project>

2. <http://www.keljob.com/editorial/actu-de-lemploi/barometre-de-lemploi.html>

3. http://data.bls.gov/timeseries/CES0000000001?output_view=net_lmth

4. <http://www.statcan.gc.ca/daily-quotidien/140307/dq140307a-fra.htm>

deux visualisations particulières du domaine, ainsi que les méthodes qui ont permis de les générer. La première visualisation est une carte dynamique permettant de savoir quels sont les métiers qui recrutent, dans quel domaine, quelle région ainsi que les compétences les plus fréquemment demandées. Construite à partir de la collection de réseaux sociaux professionnels, la seconde visualisation met en avant les parcours professionnels et permet d'observer les perspectives ainsi que les antécédents à plus ou moins long terme pour chaque métier considéré. Nous souhaitons au travers de ces visualisations aider l'étudiant en recherche d'orientation, la personne en recherche d'emploi ou en réflexion sur un changement de carrière, en lui fournissant des éléments d'information sur les perspectives d'évolution dans le métier qu'il envisage. Il pourra aussi vérifier s'il est normal qu'il n'ait pas eu de promotion dans son emploi actuel. Dans la section suivante, nous présentons des travaux liés à notre étude puis, à la section 3, les visualisations obtenues, avant de donner quelques statistiques sur les données exploitées à la section 4. Nous décrivons les différents traitements ayant permis de réaliser ces deux visualisations à la section 5.1 et à la section 5.2.

2 Travaux connexes

La plupart des travaux dans le domaine de l'e-recrutement se sont principalement intéressés à la génération de ressources linguistiques au travers de représentations, généralement dans le but d'effectuer de l'appariement de candidatures par la suite. Les travaux de le Vrang et al. (2014) présentent ainsi l'ontologie ESCO⁵, un projet européen multilingue de classification de compétences et de métiers afin de créer une harmonisation européenne en matière de recrutement. En s'appuyant sur des données extraites de réseaux sociaux ainsi que sur des offres d'emploi collectées aux Etats-Unis, Muthyala et al. (2017) proposent un moteur de recherche d'emploi qui offre à l'utilisateur des filtres concernant les informations collectées sur les sociétés telles que la taille ou encore les avantages sociaux. L'utilisation des réseaux sociaux afin d'améliorer les appariements a déjà été abordé par Diaby et Viennet (2014) qui présentent un système de recommandation afin de proposer des offres d'emploi pertinentes aux utilisateurs. Ils concluent par ailleurs que les meilleures performances sont obtenues avec la section « Compétences » de LinkedIn. Introduite en 2012 dans le cadre d'un processus de folksonomie décrit dans Bastian et al. (2014), cette section a permis aux membres de lister leurs compétences et leurs domaines d'expertise, mais aussi de recommander ou d'obtenir des recommandations de membres de leur réseau. La visualisation de données est un moyen efficace pour transmettre des informations de façon interactive et synthétique. Que ce soit au travers de nuages de mots ou au travers de projections (Blanco et Martin-Merino (2007); Sinclair et Cardew-Hall (2008)), les techniques de représentation ont souvent été utilisées pour représenter visuellement une grande variété de jeux de données. De nombreux outils ont été élaborés pour aider à la visualisation de données tels que Gephi (Bastian et al. (2009)) ou encore D3⁶ (Bostock et al. (2011)) qui permet la création de graphiques personnalisés. Ces outils sont particulièrement adaptés à la visualisation de données tout en étant des logiciels libres et flexibles. L'originalité de ce travail est donc d'exploiter les réseaux sociaux afin d'offrir des outils intelligents capables d'analyser le marché du travail et de produire une visualisation interactive et synthétique susceptible d'aider les candidats dans leur choix de carrière.

5. European Skills Competences and Occupations <https://ec.europa.eu/esco/>

6. <https://d3js.org/>

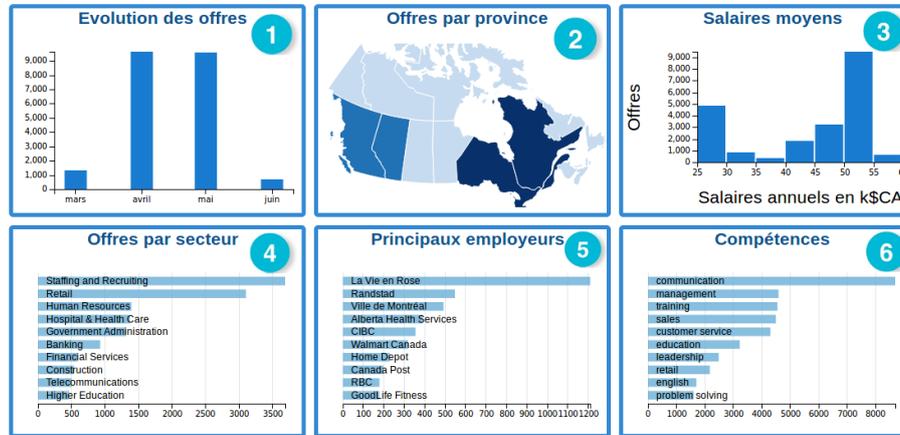


FIG. 1 – Vue d'ensemble du baromètre de l'emploi.

3 Visualisations

La figure 1 présente une vue d'ensemble du tableau de bord que nous avons désigné *Baromètre de l'emploi*. Six graphiques sont proposés en plus d'une liste déroulante permettant de filtrer en fonction d'un univers⁷. Développés à l'aide de D3, ces graphiques représentent le volume d'offres d'emploi par mois ①, la localisation géographique ②, le salaire annuel moyen ③, le secteur ④, les compagnies qui recrutent ⑤ et les compétences requises ⑥. L'utilisateur peut alors interagir avec chaque graphique, les autres se mettant automatiquement à jour. Il peut par exemple sélectionner un secteur de recrutement et connaître les principaux employeurs qui recrutent dans ce secteur d'activité, le volume d'offres d'emploi dans ce secteur.

La seconde visualisation décrit les parcours professionnels. Développée également à l'aide de D3, celle-ci permet d'afficher les parcours professionnels sous forme d'arbres de transitions. Interfaçant avec une base de données de type NoSQL du côté serveur, l'application web permet à l'utilisateur de sélectionner le métier qui l'intéresse ou d'effectuer une recherche en saisissant un métier. La figure 2 présente le résultat obtenu pour le métier *comptable* en choisissant de présenter les résultats sous forme de chronologie. Chaque transition vers un autre poste est représentée par un arc dont l'épaisseur du trait est fonction de la fréquence de la transition. Les 8 transitions les plus fréquentes sont affichées de la plus fréquente à la moins fréquente, de haut en bas. En sélectionnant un métier, l'utilisateur affiche l'ensemble des transitions possibles depuis ce noeud avec la moyenne de temps nécessaire avant de passer au métier suivant. En survolant un métier avec la souris, l'infobulle indique différentes informations comme le pourcentage de profils sur l'ensemble des profils de la collection qui ont suivi cette transition, ainsi que la durée moyenne pour passer d'un poste à l'autre. L'utilisateur peut ainsi sélectionner le métier qui l'intéresse et consulter les carrières possibles pour ce métier, les étapes nécessaires pour y arriver ainsi que les possibilités d'évolution.

7. Les univers sont une modélisation des secteurs d'activité intégré au projet BPP et qui regroupe des familles de métiers assez proches (par exemple la banque et la finance sont regroupées dans un même univers).

Visualisation de carrière à l'aide des réseaux sociaux

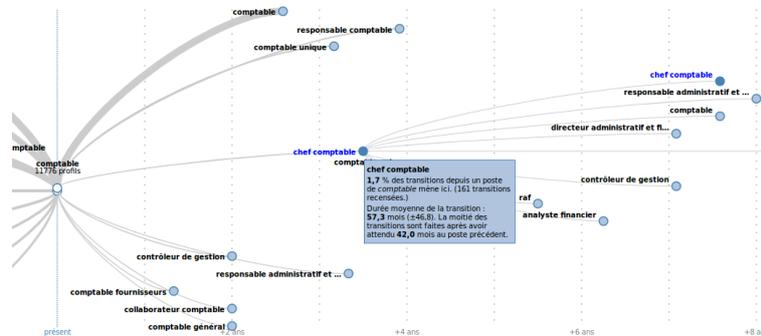


FIG. 2 – Exemple de visualisation de parcours professionnel pour le métier de comptable.

4 Données et statistiques

Nous présentons dans cette section les données qui ont permis de réaliser les visualisations. Dans le cadre du projet BPP plus de dix millions de profils issus de plusieurs réseaux sociaux (LinkedIn, Viadeo, Indeed et d'autres) ont été récoltés à l'aide d'un processus de collecte automatique. Ces données issues de profils publics professionnels ont été préalablement anonymisées puis agrégées. L'origine géographique étant le Canada (2,7 M de profils) et la France (7,5 M de profils), les profils sont soit en français, soit en anglais ou encore bilingues. Chaque profil résume différentes informations sur le parcours du candidat telles que ses diplômes et ses formations et une section du profil rend compte de ses expériences. La collection contient en moyenne 2,75 expériences par profil. Chacune contient plusieurs éléments tels que les dates de début et de fin, le nom de la société employeur, la fonction occupée par le candidat au cours de cette expérience, le lieu et un éventuel descriptif de sa mission au sein de cette société. Chaque profil regroupe un grand nombre de champs, mais il existe cependant environ 12 % de profils vides et 33 % de profils contenant peu d'information. En complément de ces données, 300 000 offres d'emploi ont été collectées sur Internet. Ces offres d'emploi couvrent un grand nombre de métiers et sont issues, elles aussi, du Canada ou de la France. Chaque offre d'emploi contient un titre, une description contenant le détail de l'offre d'emploi, la date de mise en ligne, le lieu de l'emploi proposé ainsi que le nom de la compagnie qui recrute.

5 Méthodologie

5.1 Baromètre de l'emploi

Nous souhaitons au travers de cette visualisation offrir une vue globale des tendances du marché de l'emploi. Pour ce faire, nous avons considéré chacune des composantes d'une offre à savoir le nom de la compagnie, sa localisation géographique, la date de publication de l'offre, la fonction et les compétences requises pour cette fonction ainsi que le salaire. Compte tenu de l'importance de l'aspect géographique pour cette visualisation, nous nous sommes concentrés uniquement sur les offres d'emploi canadiennes écrites en anglais, issues de la collection présentée en section 4. Identifier le nom de la compagnie et la date de publication d'une offre est

relativement facile. Géolocaliser l'offre représente un niveau de difficulté accrue. L'extraction de la fonction ou encore des compétences dans le contenu textuel d'une offre d'emploi est une tâche complexe, comme le soulignent Kessler et al. (2008). Afin d'extraire la fonction depuis le titre des offres d'emploi, nous avons ainsi constitué une liste de noms de métiers normalisés regroupant sous un terme unique les différentes écritures pour chaque métier (féminin, pluriel, erreurs typographiques, etc.) classées selon leur fréquence d'apparition. À l'aide de règles, un processus normalise et compare les titres avec cette liste de métiers. Ce processus est décrit plus en détail dans Kessler et Lapalme (2017). L'extraction automatique de compétences ayant eu de faibles résultats (F-score de 0,42), nous avons opté pour une liste de 6000 compétences, extraite des travaux de Bastian et al. (2014). Nous identifions à l'aide de cette liste un grand nombre de compétences. Plusieurs prototypes ont aussi été développés afin d'extraire le salaire, cependant les premières observations ont montré que celui-ci n'était présent que dans 14% des offres d'emploi et qu'il pouvait être formulé de nombreuses façons (taux horaire, mensuel, bonus, etc.). Afin d'offrir une vision des salaires dans le baromètre, nous avons plutôt collecté les salaires annuels moyens sur le site Emploi Québec⁸.

5.2 Parcours professionnels

Divers choix ont été faits afin de pouvoir représenter les parcours professionnels. Sur les 10 millions de profils issus de cette collection, nous recensons 26 millions d'expériences dont 23 millions contiennent une date de début et une date de fin. Pour les besoins de ce travail, nous considérons qu'il y a une transition entre deux expériences lorsque la période entre la première expérience et la seconde est inférieure à 3 mois sans chevauchement entre les deux expériences de plus de 3 mois. Nous identifions par la suite les fonctions à l'aide du processus décrit en section 5.1, ainsi que les sociétés où se sont déroulées les expériences. Nous observons une grande variation d'écriture dans les noms de compagnies tels qu'on les trouve écrits dans les réseaux sociaux, ce qui a rendu la tâche plus complexe (ex. : Mac Donalds, Mac Do, etc.). Lorsqu'un utilisateur cherche un poste, nous trouvons tous les profils contenant ce poste, puis agrégeons les parcours de ces profils. Nous recensons ainsi pour ce poste les différentes expériences précédentes et suivantes. La dernière étape consiste à transformer les résultats obtenus en arbre de transitions tel que montré à la section 3. Le tout prend moins d'une seconde.

6 Conclusion et travaux futurs

Nous avons présenté deux interfaces originales de visualisation dans le domaine de l'e-recrutement ainsi que les méthodes d'analyse de corpus pour les générer. Nous souhaitons au travers de ces outils proposer aux personnes en recherche d'emploi, en réflexion sur leur carrière, un tableau de bord synthétique et clair du marché ainsi que des perspectives professionnelles en fonction de chaque métier. Construite à partir d'offres d'emploi récoltées sur Internet, la première visualisation est une carte dynamique permettant de connaître les tendances du marché de l'emploi. Nous prévoyons d'adapter la visualisation pour traiter les offres émanant de la France et pour agréger d'autres sites d'emploi afin d'affiner les statistiques. La seconde visualisation a été constituée à partir de profils issus de réseaux sociaux, et permet

8. <http://imt.emploi.quebec.gouv.qc.ca/>

d'observer les perspectives ainsi que les antécédents pour un métier recherché. Ces travaux constituent une amorce de l'étude et de la modélisation des « parcours professionnels » qui pourraient permettre de prédire certains aspects d'expériences futures.

Références

- Bastian, M., M. Hayes, W. Vaughan, S. Shah, P. Skomoroch, H. Kim, S. Uryasev, et C. Lloyd (2014). LinkedIn skills : Large-scale topic extraction and inference. In *Proceedings of the 8th ACM Conference on Recommender Systems*, New York, NY, USA, pp. 1–8. ACM.
- Bastian, M., S. Heymann, et M. Jacomy (2009). Gephi : An Open Source Software for Exploring and Manipulating Networks. In *Int. AAAI conference on weblogs and social media*.
- Blanco, Á. et M. Martín-Merino (2007). A partially supervised metric multidimensional scaling algorithm for textual data visualization. In *IDA 2007 Springer.*, pp. 252–262.
- Bostock, M., V. Ogievetsky, et J. Heer (2011). D3 data-driven documents. Volume 17, Piscataway, NJ, USA, pp. 2301–2309. IEEE Educational Activities Department.
- Diaby, M. et E. Viennet (2014). Développement d'une application de recommandation d'offres d'emploi aux utilisateurs de Facebook et LinkedIn. In *EGC 2014*, Rennes.
- Kessler, R., N. Béchet, M. Roche, M. El-Bèze, et J. M. Torres-Moreno (2008). Automatic profiling system for ranking candidates answers in human resources. In *OTM 2008 Workshops*, pp. 625–634. Springer.
- Kessler, R. et G. Lapalme (2017). Agohra : génération d'une ontologie dans le domaine des ressources humaines. In *TAL. Volume 58 num 1/2017*, pp. 39–63.
- le Vrang, M., A. Papantoniou, E. Pauwels, P. Fannes, D. Vandenstein, et J. De Smedt (2014). Esco : Boosting job matching in europe with semantic interoperability. pp. 57–64.
- Muthyala, R., S. Wood, Y. Jin, Y. Qin, H. Gao, et A. Rai (2017). Data-driven Job Search Engine Using Skills and Company Attribute Filters. In *ArXiv e-prints*.
- Sinclair, J. et M. Cardew-Hall (2008). The folksonomy tag cloud : when is it useful? In *Journal of Information Science 34*, pp. 15–29.
- Sivabalan, L., R. Yazdanifard, et N. H. Ismail (2014). How to transform the traditional way of recruitment into online system. Volume 7, pp. 178.

Summary

In this work, we describe the analysis methods used in the creation of two original visualization interfaces in the field of e-recruitment. These interfaces were designed using millions of user profiles gathered on social networks as well as thousands of job offers collected over the internet. This work also describes the necessary steps for the implementation of the interfaces. The first visualization is a dynamic map indicating which jobs are in-demand, in which field and in which region. The second visualization tool highlights career paths and allows the user to observe the perspectives as well as the antecedents for each job considered.