Université de Montréal

Développement d'un système d'appariement pour l'e-recrutement

Par
Dieng Mamadou Alimou

Département d'informatique et de recherche opérationnelle
Faculté des arts et sciences

Mémoire présenté à la Faculté des arts et sciences

En vue de l'obtention du grade de Maitrise
en informatique

Avril, 2016

Dédicace

« Ce mémoire est dédié à ma famille : mes deux parents El Boubacar et Kadiatou Tafsir Barry qu'ils reposent en paix et à ma tendre sœur Mariama 'Hadja' qui a dû être une adulte précocement après la perte de notre maman pour pouvoir m'élever. »

Mamadou

Abstract

Our work seeks to address a very important issue in the recruitment field: matching jobs postings and candidates.

We have thousands of jobs postings and millions of profiles collected from internet provided by a specialized firm in recruitment.

Job postings and candidate profiles on professional social networks are generally intended for human readers who are recruiters and job seekers.

We use natural language processing (NLP) techniques to automatically extract relevant information in a job offer.

We use the extracted information to build automatically a query on our database.

To validate our information retrieval model of occupation, skills and experience, we use hundred Canadian jobs postings manually annotated. And to validate our matching tool we evaluate the result of the matching of ten Canadian jobs by a recruitment expert.

Keywords: job recommendation, matching, NLP, information retrieval, e-recruitment

Résumé

Ce mémoire tente de répondre à une problématique très importante dans le domaine de recrutement : l'appariement entre offre d'emploi et candidats.

Dans notre cas nous disposons de milliers d'offres d'emploi et de millions de profils ramassés sur les sites dédiés et fournis par un industriel spécialisé dans le recrutement.

Les offres d'emploi et les profils de candidats sur les réseaux sociaux professionnels sont généralement destinés à des lecteurs humains qui sont les recruteurs et les chercheurs d'emploi. Chercher à effectuer une sélection automatique de profils pour une offre d'emploi se heurte donc à certaines difficultés que nous avons cherché à résoudre dans le présent mémoire.

Nous avons utilisé des techniques de traitement automatique de la langue naturelle pour extraire automatiquement les informations pertinentes dans une offre d'emploi afin de construite une requête qui nous permettrait d'interroger notre base de données de profils.

Pour valider notre modèle d'extraction de métier, de compétences et de d'expérience, nous avons évalué ces trois différentes tâches séparément en nous basant sur une référence de cent offres d'emploi canadiennes que nous avons manuellement annotée. Et pour valider notre outil d'appariement nous avons fait évaluer le résultat de l'appariement de dix offres d'emploi canadiennes par un expert en recrutement.

Mots-clés : appariement, job recommandation, e-recrutement, big data, TALN, NLP

Remerciements

Je tiens à remercier particulièrement mon directeur de recherche Guy Lapalme pour tout son soutien mais aussi pour sa patience et sa pédagogie durant ces derniers quatorze mois qui m'ont permis d'achever ce mémoire.

Mes remerciements vont aussi à Fabrizio Gotti et Rémy Kessler pour tout leur support et leur disponibilité à répondre à toutes mes questions alors que j'étais un novice en recherche scientifique et en traitement automatique de la langue naturelle.

Merci à messieurs Antoine Gravet et Eric Tondo pour les différentes séances de travail dans leurs locaux mais surtout pour leur temps qu'ils ont voulu m'accorder afin d'évaluer mon travail mais aussi pour différentes recommandations qui ont permis en grande parti à arriver aux résultats contenus dans ce mémoire

Je ne voudrais surtout pas oublier mes collègues du RALI et du BPP qui m'ont adopté dès le premier jour au laboratoire et ont facilité mon intégration dans leur univers.

Table des matières

Liste des figures

Aucune entrée de table d'illustration n'a été trouvée.

I. Introduction

Notre époque est de plus en plus influencée par l'utilisation des données intelligentes (smart data) et du web sémantique. Le processus de recrutement n'en est pas pour autant facilité, en particulier en matière de recherche de profils et de talents, car les approches actuelles se limitent à la recherche par mots-clefs. Lorsqu'on fait une recherche par mot-clef sur des données immenses, il est évident qu'on aura de nombreux résultats parmi lesquels il faudra trier ceux qui ne correspondent pas à ce qu'on cherche. A l'ère des mégadonnées (*Big Data*), il est donc préférable de trouver d'autres approches que la recherche par mots-clés. Etant donné qu'on va manipuler des données plus ou moins bien structurées, l'une des choses envisageables serait de rechercher dans des parties spécifiques des documents ou à l'aide d'algorithmes se basant sur la recherche sémantique qui favoriseraient, par un tri pertinent, des CVs pour les recruteurs et de meilleures opportunités pour les candidats.

Problématique et motivation du TAL

On pourrait légitimement se poser la question de savoir pourquoi l'appariement d'offres d'emploi à des profils de candidats intéresse le RALI qui a pour domaine d'expertise le traitement automatique de la langue.

Comme nous manipulons dans ce projet de millions de documents textes, les techniques de NLP sont utiles pour différents problèmes :

- Extraction d'informations utiles dans du texte libre: faire ceci manuellement avec quelques documents est possible, le faire avec succès dans le cas de millions de documents nécessite plus d'expertise.
- Synonymie : un métier ou une compétence peut avoir plusieurs appellations dans la même langue (des synonymes) ce qui signifie que l'ajout des synonymes aux informations extraites peut nous permettre de trouver plus de candidats dans notre base de profils candidats.
- Traduction : une offre d'emploi est souvent dans une seule langue alors que les millions de profils que nous manipulons sont répartis entre l'anglais et le français. Plusieurs candidats sont bilingues et pourraient être intéressés par une offre d'emploi dans une autre

langue que celle de leur profil. Il est donc indispensable pour nous de traduire les informations extraites de chaque offre d'emploi dans la langue seconde sinon, on laisserait de côté des millions de profils de candidats dont le contenu textuel est dans la langue seconde. Nous pourrions utiliser les techniques de traduction automatique pour résoudre ce problème.

I.1 Contexte du projet

Le Butterfly Predictive Project (BPP) est une collaboration entre le RALI à l'Université de Montréal et LittleBigJob (LBJ), une entreprise canadienne qui œuvre depuis 2013 dans le domaine du placement de cadres en entreprise. L'activité de recrutement en ligne (e-recrutement) de LBJ se manifeste par le site `LittleBIGJob.com`. C'est une plate-forme internationale de recrutement pour les cadres et les gestionnaires. LBJ cherche à devenir le premier agrégateur d'informations en temps réel sur les profils pour permettre un appariement et une évaluation dynamique du potentiel de réussite du candidat dans un poste. LBJ souhaite améliorer considérablement la mise en correspondance entre candidats et offres d'emplois, en s'appuyant sur des outils avancés de traitement de l'information (web sémantique) et plus particulièrement par l'exploitation des données prédictives en ressources humaines (*Big Data RH*) pour l'identification de candidats intéressants mais dits passifs, car ils sont déjà en poste, mais qui pourraient être intéressés par de nouvelles opportunités.

I.2 Présentation de LittleBIGJob

LBJ offre ses services à des employeurs qui peuvent utiliser leur portail pour rechercher des profils qui les intéresseraient parmi plus de cinq millions au Canada et près de neuf millions en France.

LBJ offre aussi ses services à des cadres en recherche de meilleures opportunités de carrière en mettant en leur disposition un moteur de recherche sur toutes les offres d'emploi publiées sur internet et un service de mise en relation directe auprès des décideurs des entreprises de leur choix.

Portail

Le portail de LBJ (`littlebigjob.com`) est une solution innovante de recrutement en ligne proposant à ses clients recruteurs de :

- rechercher par mots-clefs parmi plus de 5 millions de profils;
- créer et sauvegarder un poste (offre d'emploi);
- créer et sauvegarder une liste courte *(shortlist*) de candidats pour chaque offre d'emploi;
- obtenir les coordonnées des candidats sélectionnés pour pouvoir entrer en contact avec eux.

Personal Review

Le *Personal Review* est un véritable outil d'expression de l'identité professionnelle des candidats. Il regroupe l'ensemble des éléments relatifs à leur parcours (expériences, niveau de responsabilité, motif de départ, etc.) et compétences. Il va bien plus loin qu'un simple CV et une lettre de motivation.

Career Box

LBJ permet aux candidats de conserver et protéger leurs documents professionnels dans leur Career Box. Ces fichiers peuvent être tout document relatif à leur vie professionnelle : un CV, un graphique, une présentation, un diplôme, une certification, un lien vidéo vers une conférence à laquelle ils ont assisté, etc. Cet outil complète ainsi les supports traditionnels de recrutement.

Career Manager

Un recruteur peut mandater un *Career Manager* (CMG) LBJ pour l'aider à combler un poste donné, celui-ci effectue alors des recherches avancées, puis prépare pour le recruteur une liste courte des meilleurs candidats.

En utilisant le portail de LBJ, les CMG génèrent une liste courte de candidats pour une offre d'emploi donnée. Cette liste est le résultat d'un processus de sélection qui peut se résumer par les étapes suivantes :

- Le CMG extrait manuellement des mots-clés dans l'offre d'emploi. Ces mots-clés sont : le métier ou titre du poste, les compétences, l'expérience, le salaire, etc.
- Une fois cette liste de mots-clés établie, il utilise le moteur de recherche SOLR pour trouver les candidats qui ont renseigné ces mots-clés dans leurs profils.
- Une liste courte des candidats ayant les meilleurs scores de correspondance est ainsi générée par le CMG.

Cette méthode est déjà opérationnelle et fait appel à l'expertise et l'intelligence d'un être humain durant tout le processus qui peut naturellement tenir compte de beaucoup de paradigmes propres aux recruteurs.

L'ensemble du processus étant manuel, il est relativement lent et son coût est très élevé . Automatiser l'ensemble de ce processus rendrait possible la sélection de candidats pour des milliers d'offres d'emploi en quelques secondes.

C'est ce que nous tenterons de faire dans notre mémoire et nous utiliserons le travail des CMG comme référence pour valider nos résultats.

I.3 Objectif

BPP s'appuie sur les analyses effectuées dans le cadre d'une première collaboration entre le RALI et LBJ (subvention CRSNG-Engage) qui a permis d'apprécier le potentiel de l'utilisation des méga-données (*Big Data*) dans le contexte du e-recrutement et d'effectuer des premières expérimentations sur près de deux millions de profils et offres d'emploi recueillis par LBJ sur les réseaux sociaux et sites d'emploi.

Dans le domaine des ressources humaines, choisir le candidat idéal parmi plusieurs millions de profils pour un poste est une tâche coûteuse en temps et argent. De nos jours, cette tache relève du seul bon sens d'un humain, elle se révèle pénible voire impossible dans certains cas.

Afin d'obtenir de meilleurs résultats de prédiction de la réussite d'un candidat à un poste, nous proposons d'utiliser des techniques d'appariement qui permettent de faire correspondre une offre d'emploi à une liste de candidats passifs (déjà en poste et non en recherche d'emploi) et

actifs (en recherche d'emploi) afin de d'avoir une liste courte de candidats intéressants pour ce poste.

L'objectif du projet est d'aider les *career managers* (CMG) qui sont les experts en recrutement chez LBJ dans leur travail de sélection de candidats. Actuellement les *career managers* utilisent le portail de LBJ pour trouver les profils idéals pour une offre. Cette recherche se fait par mots-clés et comporte donc beaucoup de tâches manuelles sans toutefois avoir la preuve qu'on a sélectionné les meilleurs profils pour l'offre. On essayera donc d'automatiser au mieux ce travail en s'assurant d'apparier les meilleurs profils à chaque offre d'emploi.

Nous détaillerons plus loin la procédure de sélection utilisée par les *career managers* et expliquerons comment nous avons utilisé leur travail comme référence pour la validation de nos résultats.

L'évaluation de nos résultats par monsieur Antoine Gravet qui est un CMG chez LBJ confirme que l'objectif que nous nous sommes assigné au début du mémoire a été atteint avec quelques recommandations pouvant améliorer notre outil.

II. État de l'art d'appariement d'offres d'emploi

De nombreuses études ont été faites sur l'appariement dans le domaine du e-recrutement en utilisant les données issues des réseaux sociaux. Nous étudions ces systèmes afin de proposer une solution pour notre partenaire industriel. On s'intéresse particulièrement aux systèmes de recommandation et aux ontologies dans le domaine des ressources humaines.

II.1 Les systèmes de recommandation

Nous cherchons à identifier les meilleurs profils pour une offre d'emploi. Il nous faut donc comprendre le fonctionnement des systèmes de recommandation des items à des utilisateurs car apparier une offre d'emploi avec un profil s'apparente à faire de même entre un item et un consommateur.

Selon De Campos et al. (2010), les systèmes de recommandation utilisent les préférences des utilisateurs et essayent d'en apprendre d'avantage afin d'anticiper leurs besoins.

Les utilisateurs d'internet laissent, de façon intentionnelle ou pas, des traces de leur passage qui peuvent constituer une base pour recommander aux internautes des produits qu'ils sont susceptibles d'apprécier.

Les données que nous utilisons sont des informations professionnelles publiques issues des réseaux sociaux. Elles ont été anonymisées par LBJ, nous n'avons accès qu'à un identificateur numérique. Chaque utilisateur de ces réseaux professionnels renseigne des champs qui permettent à toute personne qui a accès à leur profil de connaître leur parcours scolaire, leurs compétences, leur parcours professionnel ainsi que leur emplacement géographique.

Plusieurs approches existent pour développer un outil de recommandation. Certaines sont basées sur la recherche d'information comme celles de Baeza-Yates and Ribeiro-Neto (1999). Certains systèmes sont basés sur l'apprentissage machine comme le montre De Campos et al (2007) qui ont développé une approche basée sur les réseaux bayésiens.

Dans la plupart des systèmes de recommandation basés sur le contenu textuel, on utilise la méthode sacs de mots. Le contenu textuel d'un document est représenté par un vecteur de

couples (mot, poids du mot). Ce vecteur est souvent utilisé dans le calcul de similarité avec un autre document dont le contenu textuel est aussi représenté par un vecteur de couples.

De tous les travaux récents, celui qui se rapproche le plus à notre travail est le travail de Diaby and Viennet (2013) qui ont développé une application de recommandation d'offres d'emploi aux utilisateurs de *Facebook* et *LinkedIn* dans leur approche, ils considèrent le profil de chaque candidat comme un document et l'offre d'emploi comme un autre document ensuite ils utilisent la similarité cosinus de ces deux documents pour classer les meilleurs candidats. Nous utilisons une approche identique dans notre premier essai d'appariement mais la faiblesse de cette méthode comme nous l'expliquerons au chapitre IV réside dans le fait que pour une offre d'emploi donnée, en considérant tous les mots de sa description, il est difficile de faire ressortir uniquement les mots importants pour cette offre même en utilisant le *tf-idf* comme le font ces deux auteurs.

II.2 Les ontologies du domaine des RH

Les ontologies sont souvent utilisées dans le domaine de l'e-recrutement.

Dans le projet SIRE (Sémantique, Internet-Recrutement-Emploi), Loth et coll (2007), se basent sur une ontologie construite manuellement pour extraire les mots-clés dans une offre d'emploi; ces mots-clés seront ainsi utilisés pour la recherche de profils adéquats pour ce poste. Il s'agit d'un projet d'application des techniques d'extraction de l'information à des offres d'emploi publiées sur le web. Comme nous, ils estiment que certaines informations pertinentes notamment les compétences et les missions ne sont pas soigneusement ramassées. Pour le cas de nos données, ceci est dû au fait que l'ensemble des informations est dans du texte libre. Les membres du projet SIRE estiment tout autant que nous qu'une fois ces informations extraites et classées, elles pourrait aider à trouver une meilleure adéquation entre offres d'emploi et candidats (appariement) et fournir un plus large gamme de requêtes construites avec ces informations extraites

Une ontologie OWL et un méta-moteur spécifique sont décrits par Dorn et al. (2007); Dorn et Naz, (2007). Leur modèle privilégie la récolte de certaines informations importantes

(catégorie de l'emploi, lieu du travail, compétences recherchées, intervalle de salaire, etc.) sur un ensemble de sites web (Jobs.net, aftercollege.com, Directjobs.com etc.).

Le système IMPAKT de Colluci et coll (2009) combine une approche sémantique avec une ontologie et repose sur des méthodes de formalisation des raisonnements et des connaissances. Le système propose un appariement en effectuant des correspondances partielles ou complètes des compétences entre les offres d'emplois et les candidatures.

III. Ressources

Avant de présenter notre approche, nous décrivons les différents types de données dont nous disposons pour ce projet.

III.1 Candidats

Nos données contiennent des millions de profils anonymisés de candidats issus de réseaux sociaux professionnels publics en format JSON. Un profil contient des champs qui permettent de connaitre le parcours scolaire et professionnel d'une personne, ses compétences, sa localisation, ses compétences, etc.

ID	52b31a870b045119318b4567
EXPERI-ENCES	[2014 - ????] : CFO : *** Estate [2013 - ????] : CFO : *** Entertainment […] [1999 - 2000] : Senior Accountant : Shimm*** [1998 - 2000] : Staff Accountant : K***
SKILLS	- Project Management - Business Development - Operations Management - Management - Strategic Planning - […] - Computer Animation - Character Animation - Debt & Equity Financing
FOR-MATIONS	[1994 - 1998] : Laurentian University ;Université Laurentienne : [1992 - 1994] : Memorial University of Newfoundland :
CLAIM	CFO at Rise Real Estate
PITCH	CEO, PRESIDENT, CFO Operational Leadership, Growth and Innovation, Mergers and Acquisitions, Bank Financing, Negotiation Expertise […] Assurance, Leadership, Training and Team Building, Time Management and Prioritization, Problem Resolution and Decision-Making.

Tableau 1: exemple de profil d'un candidat

La première colonne du tableau 1 décrit les principales catégories d'informations présentes dans les profils. Le champ ID désigne un identifiant numérique unique du profil.
Le champ EXPERIENCES indique les différentes expériences du profil avec la période correspondante. La deuxième colonne présente un exemple d'informations pour un candidat dans lequel nous n'avons cité que quelques expériences puis anonymisé les noms des compagnies en remplaçant certaines lettres par des astérisques.
Le champ SKILLS nous indique les différentes compétences dont dispose le profil, on n'en a retenu que quelques-uns ici pour illustrer.

Statistiques sur les candidats

Le tableau ci-dessous donne une idée générale sur les profils et quelques champs.

	CANADA		FRANCE	
Nombre total de profils	2559115	100%	6564754	100%
Profils en français	461231	18%	4699092	72%
Profils en anglais	2097884	82%	1865662	28%
Profils sans Expérience	469131	18%	1482725	23%
Profils sans formation	948700	37%	926521	14%
Profils sans Skills	1559429	61%	4295141	65%

Tableau 2 : Statistiques sur les profils par pays

Nous avons donc la chance d'avoir une importante base de profils candidats pour nos expériences. Même si plusieurs profils sont vides ou pauvres en contenu (61% sans compétences renseignées) mais il en reste tout de même un grand nombre pour expérimenter avec nos méthodes d'appariement.

III.2 Offres

On dispose aussi d'offres d'emploi récoltées par notre partenaire industriel, ces offres sont encodées en format JSON. Dans une offre d'emploi, plusieurs informations nous permettront de l'apparier avec des profils de candidat. On retrouve dans chaque offre une

description du poste, son emplacement géographique, la mission du futur détenteur du poste, les exigences en termes d'éducation, d'expérience professionnelle et de compétences.

Nous travaillerons sur une base de données de plus de 60 000 offres d'emplois canadiennes.

La figure 3 représente un exemple d'offre d'emploi.

ID	5553678103be984ac58019b8
TITRE	Maintenance Millwright
DESCRIP-TION	**** Industries Limited has enjoyed steady growth over the years providing metal finishing solutions. To continue our growth we are currently seeking a dynamic, well-organized Maintenance Technician/Millwright to join our team. SUMMARY: The primary duties of the Maintenance Technician/Millwright is to maintain and repair plant equipment MAJOR RESPONSIBILITIES/DUTIES: Surveillance of plant equipment and shop upkeep • Organize and prioritize workload • Ensure minimal down-time of equipment […] • Use hand and power tools and welding equipment Health & Safety • Ensure Compliance REQUIREMENTS / QUALIFICATIONS: • Educational background in mechanical engineering, coupled with a Millwright and/or Electrician license (or equivalent) is considered an asset • Minimum of five (5) years experience working in a related position • Excellent interpersonal, written and verbal communication skills • Solid trouble-shooting and problem solving skills • Ability to prioritize, plan and schedule work effectively with a pro-duction team in a fast paced and values driven environment Strong organization and time management skills to ensure deadlines are met.
PLACE	Cambridge, ON
COMPANY NAME	**** Industries

Tableau 3 : exemple d'une offre d'emploi

Le champ ID est un identifiant numérique de l'offre. Le champ DESCRIPTION nous donne des informations sur l'entreprise et le poste. On y retrouve les compétences et l'expérience exigées mais non balisées. Le champ PLACE renseigne l'emplacement géographique de l'offre d'emploi tandis que le champ COMPANY NAME indique le nom de la compagnie.
La plus grande difficulté réside dans l'extraction automatique des informations dans le champ DESCRIPTION, ceci nécessite de combiner plusieurs techniques de traitement automatique de la langue.

III.3 Ontologie ESCO

ESCO le Vrang, Papantoniou et al. 2014 est une ontologie du domaine des ressources humaines qui a été développée par une agence de l'union européenne. Il s'agit d'une classification des compétences, qualifications et occupations européennes. Cette ontologie a l'avantage d'être multilingue, mais nous nous limitons aux appellations en anglais et en français. Elle propose des appellations alternatives pour chaque compétence, qualification et occupation. En outre, il existe des liens entre les occupations et les compétences de sorte que dans ESCO on peut retrouver les compétences de chaque métier ou groupe de métiers. Nous décrirons plus loin comment nous nous en sommes servis pour extraire les informations utiles dans les offres d'emploi.

III.4 CNP-NOC

La classification nationale des professions canadiennes est une ressource gouvernementale canadienne qui répertorie tous les métiers qu'on rencontre dans l'économie canadienne. Comme tout document fédéral officiel, elle est bilingue ce qui est un avantage. Comme l'ontologie ESCO, CNP-NOC nous sera très utile dans l'extraction des items utiles contenus dans les offres d'emplois.

III.5 Elite20

LBJ nous a fourni une référence de 64 offres d'emploi avec pour chaque offre, vingt candidats sélectionnés par un CMG. On a créé une arborescence équivalente à notre base de données MySQL dans laquelle pour chaque job, on a un répertoire qui porte le nom de son ID et on

retrouve dans ce répertoire un fichier JSON de l'offre d'emploi et un autre contenant les profils des candidats retenus pour cette offre.

III.6 Dictionnaire bilingue de métiers

Dans le cadre de ce mémoire il nous est arrivé de produire des ressources pour les utiliser mais aussi pour les mettre au service de la communauté scientifique. L'une de ces ressources est le dictionnaire bilingue des métiers.

Pour construire ce dictionnaire, nous avons utilisé deux ressources :

- ESCO : ontologie, décrite à la section III.3, compte 5381 occupations et pour chaque occupation, on peut avoir jusqu'à quatre appellations alternatives dont deux en français ou en anglais.
- CNP-NOC (section III.3) de la classification nationale des professions, on en extrait 36 540 occupations filtrées pour n'avoir que des métiers. Une traduction en langue seconde est disponible pour chaque occupation.

Le dictionnaire bilingue que nous avons constitué est la fusion de ces deux ressources en prenant soins de faire certains prétraitements. La ressource ainsi constituée est disponible sur le site web du projet.

III.7 Référence de cent offres d'emploi annotées manuellement

Dans le prochain chapitre, nous allons procéder à l'extraction automatique de certaines informations dans les offres d'emploi. Pour évaluer nos extracteurs automatiques, nous devons utiliser une référence pour vérifier la performance de nos algorithmes.

Pour établir cette ressource qui va nous servir de référence, nous avons tiré au hasard cent offres d'emploi parmi six mille limitées aux métiers qui intéressent LBJ.

Pour ces offres, nous avons extrait manuellement le titre, le nombre d'années d'expérience et les compétences.

Le titre : Le titre d'une offre d'emploi est toujours renseigné dans le champ qui porte le même nom. Mais dans la plupart ces cas, certaines informations sont ajoutées par l'employeur pour apporter plus de précision sur l'offre d'emploi. Ces informations peuvent être le secteur d'activité de l'entreprise (construction, aéronautique…), le statut de l'emploi (temps plein, temps partiel, temporaire, permanent), la localisation du poste (ville, province, pays...) ou toute autre information que l'employeur a jugé utile.

Cependant, on entend par le titre d'une offre d'emploi, le métier. C'est pourquoi, on n'annote ici que le métier ou l'occupation contenus dans le titre de l'offre d'emploi.

Dans le Tableau 4, nous présentons quelques exemples de titres annotés.

JOB-ID	TITLE ORIGINAL	TITRE ANNOTE
55827d36acdfa2adb63b5cd5	sales /account executive	sales executive
5552997103be984ac5801568	assistant store manager - guelph	assistant store manager
55390903c51c771c4bc4978a	account manager - montreal	account manager
5516b674c51c771c4bc37f83	territory manager	territory manager
554d8b4403be984ac580010d	assistant buyer regional merchandising	assistant buyer
5547feaf03be984ac57fe569	sales professional - owner direct	sales professional
555e765803be984ac5804a05	commercial leasing manager	commercial leasing manager
55427bbd03be984ac57fd0da	production manager	production manager
5533a239c51c771c4bc43a32	senior financial analyst johnson & johnson medical companies markham toronto on job	financial analyst
5548bf4a03be984ac57fe8d9	commercial construction superintendent	commercial construction superintendent
5565a97703be984ac580661a	sales representative - ats technology recruiting and screening services toronto 165551	sales representative
5553501203be984ac580194a	account manager (telecommunications)	account manager
5553004f03be984ac58016ec	senior financial analyst sales opex	financial analyst
55178818c51c771c4bc38248	methods agent - seat - global 7000/8000	methods agent
551e6e53c51c771c4bc38e4e	territory manager alberta south (nutro)	territory manager
555afc0703be984ac580354f	senior financial analyst	financial analyst
55454db403be984ac57fdb27	inside sales manager â€“ cloud human capital management technology	sales manager
5553c57403be984ac5801c18	hourly payroll administrator	hourly payroll administrator
5548bef803be984ac57fe8cd	accounting clerk	accounting clerk
553a6d6ac51c771c4bc4c290	avionic methods agent	avionic methods agent
554f3b0e03be984ac580050b	sales account manager	sales account manager

Tableau 4: annotation du titre d'offres d'emploi

Expérience : On entend par expérience, le nombre d'années requises pour le poste. On retrouve dans la majorité des cas, une phrase explicite dans le champ DESCRIPTION qui nous renseigne sur l'expérience**.**

Le tableau 5 suivant illustre quelques expériences annotées.JOB-ID	DESCRIPTION	EXPE-RIENCE
55827d36acdfa2adb63b5cd5	A full time Sales/ Account Executive is being sought by an exciting Canadian company ... Qualifications 1-2 years sales experience preferably B2B though…..	1
5552997103be984ac5801568	Synergie Hunt International is currently seeking high energy retail managers..... Job Requirements Minimum of 5 years of relevant experience in the retail industry including at least 2 years in a supervisory position….	5
55390903c51c771c4bc4978a	Sales Account Manager - Mobile Accessories Cesium.... Required experience: Sales!!!!!: 3 years	3
5516b674c51c771c4bc37f83	Our client is a global distributer of medical aesthetics and currently they have asked Lock Search.... Bachelor's degree or equivalent 2+ years' of relevant sales experience and/or device sales experience is considered an asset...	2
551658c7c51c771c4bc37dd5	This position is responsible to assist in the promotion of grain programs and services.... Minimum of three years of plant experience including receiving grading and customer service….	3

Tableau 5: annotation de l'expérience d'offre d'emploi

Compétences : On retrouve les compétences requises pour le poste dans le champ description de l'offre d'emploi. Cependant, elles ne sont pas explicitement citées comme on le verra au Tableau 6 qui illustre l'annotation des compétences d'une offre d'emploi :

ID	5553004f03be984ac58016ec
TITRE	methods agent
DESCRIP-TION	Methods Agent…. Analyze the feasibility of customer requests (P&O C&O) and program change requests (PCR) Plan and implement changes in production (all change affecting the production method) and coordinate first unit in production. Support daily production activities Identify non-value added activities eliminate them and optimize production Establish the following requirements: ergonomics & health & safety tools and production equipment according to optimal work methodology. Participate in the development and maintenance of methods procedures. Participate in continuous improvement through the achieving excellence system. Qualifications As our ideal candidate You have a technical college degree and/or equivalent experience You have a minimum of 5 years of experience in aerospace or other related manufacturing industry. Experience in the upholstery of seat structures will be considered an asset. You are organized a good planner and problem solver You have a good judgment have a sense for innovation and possess a good decisional process You have good interpersonal communication skills and are customer oriented You have strong project management skills and function well within a multi-disciplinary team You are able to read technical drawings and have good visualization skills in 3D. You have good knowledge of MSOffice suite (Word Excel Powerpoint Outlook) You are functional in the French language spoken and written Knowledge of aircraft structures systems or interiors will be considered an asset Knowledge of SAP and CATIA V5 will be considered an asset….
COMPÉ-TENCES ANNOTÉES	organized, good planner, problem solver, sense for innovation, good interpersonal communication skills, customer oriented, project management, visualization skills, ms office, Word, Excel, Powerpoint, Outlook, French, aircraft structures, SAP, CATIA V5,

Tableau 6: exemple d'annotation de compétence

L’annotation manuelle de ces cent offres d’emploi constitue une ressource importante dans le cadre du projet. Nous allons l’utiliser dans le Chapitre IV pour évaluer nos algorithmes d’extraction d’informations (titre, expérience et compétences). La ressource est aussi mise à la disposition des autres intervenants des projets.

IV. Appariement

Dans ce chapitre, nous décrivons deux approches à l'appariement que nous avons développées.

IV.1 WordMatch

WordMatch est notre premier algorithme de correspondance entre une offre d'emploi décrite à la section III.2 et une liste de candidats (section III.1).

Description

L'offre d'emploi contient deux champs importants décrivant les exigences du poste notamment le champ DESCRIPTION. Nous avons constitué un sac de mots avec ces deux champs, puis nettoyé cet ensemble en éliminant les mots les plus courants de la langue (anglais ou français dépendamment de la langue de l'offre).

Nous constituons aussi une liste de mots à partir des champs SKILLS, EXPERIENCE et EDUCATION du candidat puis on la nettoie de la même façon que pour l'offre c'est-à-dire en éliminant les mots les plus courants de la langue. Une fois les deux listes constituées et nettoyées, nous les comparons et retournons leur intersection. Nous désignons par score d'un candidat, la longueur de son intersection avec l'offre. Nous pouvons ainsi classer les candidats par ordre décroissant de score.

Évaluation

Après avoir vérifié à la main les profils des candidats ayant les plus grands scores pour quelques offres, nous avons remarqué que la plupart des candidats répondaient plus ou moins à quelques critères de l'offre.

Cependant, les résultats ont été analysés par un expert CMG de chez LBJ et comportaient plusieurs incohérences entre l'offre d'emploi et les candidats sélectionnés. Ceci est dû au fait que nous recherchons un ensemble de mots qui peuvent figurer parfois dans des profils qui ne nous intéressent pas pour cette offre donnée.

Si les items que forment nos sacs de mots étaient pondérés selon leur importance pour un type ou groupe de document spécifique, ceci permettrait d'apparier pour correctement en utilisant cette méthode.

Par ailleurs, en essayant de retrouver les candidats ELITE20 de notre référence, le rappel était proche de zéro. Car le CMG utilise certains procédés (propres aux recruteurs), tels la traduction, la synonymie… il cherche des mots-clés différents de ceux dans l'offre d'emploi ou dans une autre langue.

Les résultats de cette première approche n'ayant pas donné les résultats escomptés, nous avons décidé d'en développer une deuxième qui intègre certains paradigmes propres aux recruteurs.

IV.2 SkillFinder

Les différentes méthodes que nous avons rencontrées dans la littérature proposent d'extraire les mots-clés dans les offres d'emploi et les profils de candidats en utilisant une pondération, le plus souvent le *tf-idf,* pour définir l'importance d'un mot par rapport à un autre. Ensuite ces auteurs utilisent la similarité cosinus Mamadou et Viennet 2013 pour les apparier.

Nous utilisons nos ressources pour l'extraction, la traduction, et l'extension des informations dans les offres d'emploi. L'indexation sera utilisée pour la recherche des profils correspondant aux exigences de chaque offre d'emploi.

La phase la plus longue et la plus importante de notre approche est l'extraction des informations utiles dans les offres d'emploi. L'extraction du métier, des compétences et de l'expérience est détaillée dans les sections suivantes.

La figure 6 illustre notre processus d'appariement.

Les références dans la figure sont les numéros de sections dans lesquelles sont détaillées les étapes y référant.

On part d'une offre d'emploi qu'on veut apparier avec notre banque de profils. La première étape consiste à extraire les informations pertinentes dans l'offre d'emploi :

- le métier contenu dans le titre de l'offre d'emploi en effectuant des traitements qui sont détaillés dans IV.2.1.
- les compétences contenues dans la description de l'offre d'emploi en suivant notre méthodologie que nous expliquons dans IV.2.2.
- le nombre d'années d'expériences exigées par l'offre d'emploi, la méthode d'extraction est détaillée dans IV.2.3.

Après avoir extrait le métier et les compétences, on interroge notre banque de profils pour sélectionner les candidats qui ont déjà occupé le métier extrait similaire et qui ont les compétences extraites; l'interrogation est détaillée dans IV.2.4.

Les candidats filtrés par interrogation sur le métier et les compétences sont ensuite filtrés par l'expérience extraite et on génère ainsi une liste courte de candidats.

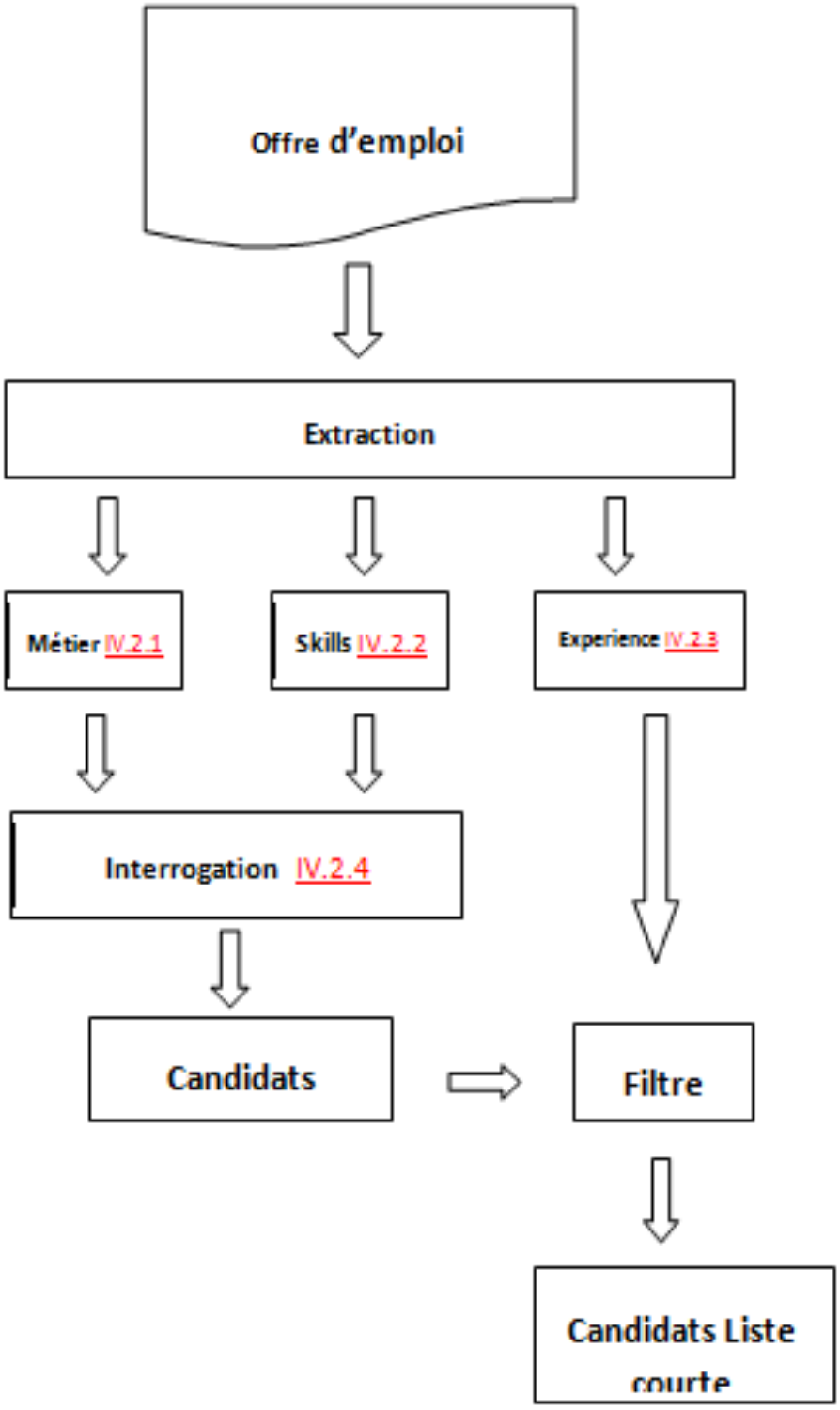

Figure 6 : processus d'appariement SkillFinder

IV.2.1 Extraction du métier

Le métier d'une offre d'emploi étant renseigné dans le champ TITLE, on arrive à l'extraire facilement. Comme nous l'avons constaté lors de notre annotation manuelle (III.7), la problématique réside dans son prétraitement car dans beaucoup d'offres, les employeurs rajoutent des informations telles le statut de l'emploi, l'horaire, l'emplacement géographique du poste, le secteur d'activités, etc.

L'extraction du métier se fait en éliminant ce bruit par les étapes suivantes :

- suppression du secteur d'activité : on a identifié des séparateurs qui permettent dans plus de 80% des cas d'éliminer le secteur d'activité.

  Exemple de titre : directeur marketing – publicité.

  Dans ce titre publicité indique le secteur d'activité de l'offre d'emploi
- suppression du bruit : en éliminant des mots nous avons constitué manuellement un antidictionnaire pour cela. Exemples de bruit : *temps plein, 12$/h, poste basé a Montréal etc…*
- On vérifie si un des métiers les plus fréquent du dictionnaire des métiers III.6 des métiers fait partie du champ titre. Si oui, on considère ce métier.

Exemple d'extraction de métier

En considérant l'exemple d'offre d'emploi illustré au Tableau 3 dans la section des ressources, nous avons appliqué notre modèle d'extraction de métier décrit dans IV.2.1 et le résultat est présenté au Tableau 7.

ID	5553678103be984ac58019b8
TITRE	Maintenance Millwright
TITRE EXTRAIT	Maintenance Millwright

Tableau 7: exemple d'extraction de métier

Évaluation

Pour évaluer cette tâche, nous avons appliqué notre méthode d'extraction aux cent offres d'emploi qui constituent notre référence et nous avons mesuré la précision et le rappel :

Précision	0.89

Rappel	0.89

Tableau 8: évaluation de l'extraction du titre

On remarque que pour cette tâche, la précision est égale au rappel car dans tous les cas (pour les cent offres d'emploi), on extrait un seul titre (métier).

IV.2.2 Extraction des compétences

L'extraction des compétences dans les offres est une tâche beaucoup plus difficile que celui du métier. La difficulté est d'extraire les compétences minimales dans du texte libre. Pour pouvoir extraire de ce texte seulement les groupes de mots qui sont des compétences, on procède comme suit :

- On construit un arbre de compétences avec la liste de compétences les plus fréquentes dans les profils.
- On recherche dans cet arbre les suite de mots du champ DESCRIPTION des offres dans l'arbre des compétences. La liste générée ne comporte que des compétences à moins que le dictionnaire à l'aide duquel on a construit l'arbre ne comporte du bruit.

Exemple d'extraction de compétences

En considérant l'exemple d'offre d'emploi illustré par le Tableau 3 dans la section des ressources, nous avons appliqué notre modèle d'extraction des compétences décrit dans IV.2.2 Le tableau 9 illustre un exemple d'extraction de compétences.

ID	5553678103be984ac58019b8
TITRE	Maintenance Millwright
COMPÉTENCES EXTRAITES	Repair, industrial, pneumatic, including, drawings, logic, driven, mechanical, engineering, organization and time management skills, hydraulic, equipment, Electrical, maintenance, power tools, team, interpersonal, well-organized, grinders, corrective action, work effectively, blueprints, installation, etc.

Tableau 9: exemple extraction de compétences

On arrive à extraire certaines compétences telles *pneumatic, mechanical engineering, organization and time management skills, well organized* etc. Mais on extrait d'autres items qui ne représentent pas des compétences *: repair, including, team, equipment* etc. Ceci est dû au fait que nous avons ces items dans notre liste de compétences qui avait été construite automatiquement.

Évaluation de l'extraction des compétences

Pour évaluer cette tâche, nous avons appliqué notre méthode d'extraction de métier aux cent offres d'emploi qui constituent notre référence et nous avons mesuré la précision et le rappel. Contrairement à l'extraction du titre et de l'expérience, pour évaluer les performances de cette tâche, on calcule le rappel et la précision pour chaque offre d'emploi.

Nous avons évalué notre outil pour plusieurs seuils soit le nombre de fois qu'une compétence doit apparaitre dans nos profils candidats pour que nous la considérions comme compétence. Pour chacune des fenêtres, nous avons calculé trois métriques pour nous permettre de déterminer la meilleure fenêtre.

Les trois métriques dans le tableau 10 sont des moyennes pour les cent offres d'emploi de notre référence.

SEUILS	COMPÉTENCES EXTRAITES	PRÉCISION	RAPPEL	F-SCORE
10	27	0,16	0,52	0,23
50	18	0,24	0,55	0,31
100	14	0,32	0,53	0,36
200	11	0,38	0,51	039
300	9	0,41	0,50	0,41
400	8	0,45	0,48	0,42
500	8	0,45	0,46	0,41
600	7	0,48	0,44	0,42
700	7	0,48	0,43	0,41
800	6	0,48	0,41	0,40
900	6	0,50	0,40	0,40
1000	6	0,52	0,40	0,40
1500	5	0,52	0,37	0,39
2000	4	0,55	0,32	0,36

Tableau 10: évaluation de l'extraction des compétences en fonction de la liste des compétences

En observant les résultats du Tableau 10, nous avons décidé de fixer notre fenêtre à 1000. Cette fenêtre nous permet d'avoir une meilleure précision tout en extrayant un nombre de compétences raisonnables pour notre appariement.

Le Tableau 11 présente les moyennes des résultats sur les cent offres avec une fenêtre de 1000.

Moyenne précision	0,52
Moyenne rappel	0,40
Moyenne f-score	0,40

Tableau 11: évaluation de l'extraction des compétences

IV.2.3 Extraction de l'expérience

L'expérience, le nombre d'années dans un poste similaire, est une information très importante dans une offre d'emploi. Dans nos offres d'emploi canadiennes, cette information est dans le champ DESCRIPTION qui est du texte libre. Pour réussir notre appariement, on doit pouvoir extraire l'expérience avec un taux de succès acceptable.

Voici la méthode que nous proposons pour extraire efficacement l'expérience exigée par dans une offre d'emploi :

- On construit une liste de tous les termes qui font référence à une expérience: *experience, expérience, years, années, related ….*
- On construit une expression régulière avec cette liste
- A l'aide de de notre expression régulière, on cherche l'expérience dans le champ description de nos offres d'emploi.

Exemple d'extraction d'expérience

En considérant l'exemple d'offre d'emploi illustré au Tableau 3 dans la section des ressources, nous avons appliqué notre modèle d'extraction du nombre d'années d'expériences à partir de la phrase*: Minimum of five (5) years experience working in a related position.*

Le résultat est donné dans le Tableau 12.

ID	5553678103be984ac58019b8
TITRE	Maintenance Millwright
EXPERIENCE EXTRAITE	5

Tableau 12: exemple extraction d'expérience

Évaluation de l'extraction de l'expérience

Pour évaluer cette tâche, nous avons appliqué notre méthode d'extraction de métier aux cent offres d'emploi de notre référence et nous avons mesuré la précision et le rappel.

Expériences extraites manuellement	68
Expériences extraites automatiquement	65
Expériences correctement extraites	58
Précision	0.89
Rappel	0.85

Tableau 13: évaluation de l'extraction d'expérience

.

Extension du titre

Le métier ainsi extrait peut être étendu par sa traduction en langue seconde et ses synonymes. Cette étape est possible grâce au dictionnaire bilingue de métiers que nous avons construit. Le tableau 14 illustre l'utilisation de ce dictionnaire pour l'extension du titre.

ID	5553678103be984ac58019b8
TITRE	Maintenance Millwright
TITRES ALTERNATIFS	Mécanicien de chantier à l'entretien, mécanicienne de chantier à l'entretien

Tableau 14: exemple extension du titre

IV.2.4 Recherche de profils

Dans les sections précédentes, nous avons décrit l'extraction d'informations contenues dans une offre d'emploi. Ces informations ont été extraites pour rechercher les candidats qui ont renseigné ces mêmes informations dans leurs profils.

Rappelons que nous disposons d'un index LUCENE de profils tels que décrit dans le Tableau 1. Notre index est construit de façon à interroger des champs spécifiques des profils.

Nous devons donc construire des requêtes booléennes avec les informations extraites.

Construction de requête

On construit deux requêtes principales : une avec le métier extrait de l'offre d'emploi et une autre avec les compétences extraites de la même offre.

Requête de titre

On construit la requête du titre en recherchant le titre extrait dans l'offre d'emploi dans les différentes fonctions occupées par les candidats lors de leurs expériences. On peut par exemple rechercher le métier *software programmer* dans le champ function :

`expfunc: 'software programmer'.`

Requête de compétences

Les compétences extraites dans l'offre d'emploi sont aussi recherchées dans le champ skills des candidats. Par exemple on peut rechercher la compétence *java* en utilisant :

`skillname: 'java'`

Requête booléenne

Une requête sur le métier et les compétences nous permettant de rechercher des profils est automatiquement générée par notre système.

En plus de chercher le métier dans les fonctions occupées par les candidats lors de leurs différentes expériences, on le cherche aussi dans leur présentation.

En effet, beaucoup de candidats ne renseignent pas leurs expériences dans les champs appropriés, mais décrivent leur parcours dans le *personal branding claim* ou le *personal brinding pitch* qui sont des champs de présentation.

La requête booléenne générée est de la forme :

(expfunc:"sales account executive" **OR claim:**"sales account executive" **OR pitch:**"sales account executive"**) AND skillname:**(sales)

Les profils retournés par cette requête sont alors filtrés en éliminant ceux qui n'ont pas l'expérience requise par l'offre d'emploi.

V. Évaluation de SkillFinder

Dans cette section, nous présentons les résultats de notre processus d'appariement décrit en « référence ».

Nous avons utilisé deux listes d'offres d'emploi : la liste des offres Elite20 (voir III.5) pour laquelle on a déjà pour chaque offre d'emploi une liste de candidats sélectionnés par un CMG et la liste des cent offres d'emploi canadiennes (voir III.7) qui nous a servi de référence pour évaluer nos différents modules d'extraction d'information.

V.1 Résultats sur les offres d'emploi Elite20

Pour chaque offre d'emploi Elite20, nous avons déterminé le nombre de candidats communs entre ceux retenus par notre outil et ceux sélectionnés par le CMG. Nous avons aussi calculé le temps utilisé par notre outil pour sélectionner des candidats.

Le tableau 15 présente les résultats pour les offres d'emplois pour lesquels nous avons au moins deux candidats en communs.

- La colonne TITRE indique le titre de l'offre d'emploi sans aucun traitement fait par nous.
- La colonne #CAND indique le nombre de candidats sélectionnés par notre outil.
- La colonne #CAN E20 indique le nombre de candidats sélectionnés par le CMG pour cette offre d'emploi.
- La colonne #COMMUNS indique le nombre de candidats communs entre ceux sélectionnés par le CMG et ceux sélectionnés par notre outil.
- La colonne #TEMPS indique le temps que prend notre outil pour sélectionner des candidats.

TITRE	#CAND	#CAN E20	#COMMUNS	TEMPS
directeur d'usine	135	32	10	8.25
spécialiste d'applications en génie civil	174	10	8	1.35
développeur java, serveur et interface web	181	25	7	1.54
developpeur mobile (android, ios et windows phone)	181	20	6	1.56
maitre de chai / vinificateur	148	20	5	3.23
contrôleur de projets (pco)	183	26	5	1.61
lead marketing / contenu	190	26	4	10.29
spécialiste seo	191	9	4	1.60
directeur marketing	198	27	4	4.92
gestionnaire de compte / territory account manager	187	26	3	10.75
acheteur/euse	90	12	2	4.37
directeur principal des soins	194	10	2	1.43
account director	142	14	2	7.36
senior vp sales	150	26	2	1.73
senior vp sales	150	21	2	1.68
représentant des ventes externes (pièces camions lourds)	177	17	2	1.28
technicien, campagnes électroniques crm	192	4	1	1.31
ingénieur instrumentation	146	12	1	1.30
directeur régional des ventes	196	27	1	1.43
concepteur/designer-cuisiniste	178	7	1	13.22
représentant des ventes senior	140	21	1	1.35
gestionnaire de projet ecommerce	193	20	1	1.55
directeur au développement des affaires	160	21	1	1.73
directeur(trice) de projets	146	32	1	1.50
moyenne	**167**	**19**	**3**	**3.59**

Tableau15 : résultats appariement sur offres Elite20

Bien que notre modèle sélectionne des candidats pertinents pour chaque offre d'emploi, nous avons quand même peu de candidats communs avec ceux sélectionnés par le CMG. Ceci est principalement dû au fait que dans notre modèle, on ne tient pas compte du secteur d'emploi qui est une information que nous n'extrayons pas car elle n'est pas renseignée.

Les CMG sélectionnent les candidats en prenant soin d'éliminer ceux qui ne sont pas du même secteur d'activité que l'offre d'emploi.

Les sélections du CMG reflètent une certaine subjectivité et certains paradigmes propres aux recruteurs.

V.2 Résultats sur les cent offres d'emploi canadiennes de notre référence annotée

Nous avons constaté à la section précédente que les candidats sélectionnés par notre outil sont pertinents, mais cette affirmation serait plus crédible si elle venait d'un expert en recrutement.

Ainsi, persuadés qu'il est normal que notre système n'arrive pas à sélectionner les mêmes candidats qu'un CMG, nous avons sollicité LBJ pour analyser nos résultats. Dans notre cas, c'est la précision qui nous intéresse c'est-à-dire le taux de candidats pertinents.

Monsieur Antoine Gravet, CMG chez LBJ, a accepté d'analyser la pertinence de nos candidats sélectionnés par rapport aux offres d'emploi. Nous lui avons soumis dix offres d'emploi choisies au hasard parmi les cent. Pour chacune, nous avons sélectionné jusqu'à (voir la colonne remarques dans le tableau) dix meilleurs candidats. Monsieur Gravet a jugé la pertinence de chaque candidat en ajoutant des remarques pour nous permettre de perfectionner notre outil.

Le tableau 16 montre l'évaluation de monsieur Gravet sur les 4 premières offres évaluées.

OFFRE	CANDIDATS	PERTINENCE	COMMENTAIRE
1	1	pertinent	trop sénior
	2	pertinent	trop sénior
	3	pertinent	trop sénior
	4	pertinent	trop sénior
	5	pertinent	confirmé
	6	pertinent	trop sénior
	7	non pertinent	secteur différent
	8	pertinent	trop sénior
	9	pertinent	trop sénior
	10	non pertinent	secteur différent
2	1	pertinent	trop sénior
	2	non pertinent	secteur différent
	3	pertinent	confirmé
	4	pertinent	
	5	non pertinent	
	6	pertinent	trop sénior
	7	pertinent	trop sénior
	8	pertinent	trop sénior
3	1	pertinent	confirmé
	1	pertinent	confirmé
	3	pertinent	secteur différent
	4	pertinent	trop sénior
	5	pertinent	trop sénior
	6	pertinent	trop sénior
	7	pertinent	confirmé
	8	pertinent	confirmé
	9	pertinent	secteur différent
4	1	pertinent	confirmé
	2	pertinent	confirmé
	3	pertinent	trop éloigné
	4	pertinent	confirmé
	5	pertinent	confirmé
	6	pertinent	secteur différent
	7	pertinent	secteur différent
	8	pertinent	secteur différent
	9	pertinent	secteur différent
	10	pertinent	confirmé

Tableau16 : Évaluation de l'appariement par Antoine Gravet

Le tableau 17 présente l'ensemble des résultats de l'évaluation sur les 10 offres.

candidats	nombre	pourcentage
étudiés	81	
jugés pertinents	69	85%
jugés non-pertinents	12	15%
jugés trop seniors	23	28%

Tableau 17 : Évaluation par Antoine Gravet

Les résultats de l'évaluation de notre système montrent que 85% de nos candidats sont pertinents. Pour les candidats jugés trop séniors, l'explication vient du fait que SkillFinder considère les candidats ayant au minimum le nombre d'années exigées par l'offre d'emploi sans toutefois tenir compte du critère de la durée de leur expérience. Pour corriger ceci les experts de chez LBJ ont recommandé de ne sélectionner que les candidats avec au plus trois années de plus que le nombre d'années exigées par l'offre.

L'autre recommandation de LBJ est d'intégrer dans notre processus d'appariement le secteur d'activité.

Pour extraire le secteur d'activité, deux pistes pourraient être explorées. La première chose que nous pourrions faire serait d'étudier le vocabulaire propre à chaque secteur d'activité pour ensuite classer chaque offre d'emploi dans un secteur. La seconde piste serait d'utiliser l'information sur la compagnie, cette information est renseignée dans nos offres d'emploi et d'utiliser la table de correspondance Compagnie-secteur d'activité afin de classer notre offre d'emploi dans le secteur correspondant.

V.3 Annotation du secteur d'activité :

Lors des différentes rencontres avec les experts de chez LBJ, plusieurs remarques ont porté sur la non prise en charge par notre système du secteur d'activité. Malgré le fait que notre système ne prend pas en compte le secteur d'activité, la prise en compte par notre système des compétences minimise le nombre de candidats hors secteur d'activité.

Avant de nous pencher sur l'extraction automatique du secteur d'activité nous avons voulu évaluer la valeur ajoutée que cette information pourrait apporter à notre système. Nous avons donc choisi l'offre d'emploi pour laquelle nous avions le plus de remarques indiquant que nous avons choisi des candidats hors secteur d'activité pour annoter manuellement le secteur d'activité pour ensuite reprendre le processus d'appariement avec skillFinder afin d'évaluer le résultat.

Considérons l'offre 4 des offres d'emploi canadiennes annotées; voir Tableau 16 pour les remarques. Le tableau 18 illustre cette offre.

ID	5516b674c51c771c4bc37f83
TITLE	territory manager
DESCRIPTION	**OVERVIEW:** As the Territory Manager you will report to the Manager of Regional Sales and work closely and under the direction of the Sales Manager. You will assist in meeting assigned territory revenue and objectives along with satisfying customer needs. **RESPONSIBILITIES:** Bachelor's degree or equivalent 2+ years' of relevant sales experience and/or device sales experience is considered an asset. Possess the ability to be self-sufficient while also taking direction accordingly. The ability to show confidence in the field while also willing to receive constructive feedback simultaneously. Successful record of achievement with a demonstrated history of initiative and achievement Proven successful sales track record required. Excellent negotiation, communication, and organizational skills. Travel is required. **RENUMERATION:** Competitive Salary (100K plus T4) + Benefits + Bonus + Vehicle Allowance.

Tableau 18 : offre d'emploi à annoter pour le secteur d'activité

Pour l'annotation du secteur d'activité, nous utilisons un tableau de secteur d'activité fourni par LBJ.

ID Univers	Description
1	Administration publique, territoriale et internationale
2	Aéronautique et aérospatiale
3	Agriculture, viticulture, élevage, pêche
4	Agroalimentaire
5	Art, design, culture et artisanat d'art, musique, musée
6	Associations, syndicats, fondation, social, humanitaire, religions
7	Assurances, mutuelles, prévoyance
8	Audiovisuel, cinéma, spectacles, média, publicité, événementiel, divertissement, com
9	Industries automobiles
10	Banque, finance, capital risque, fonds privés
11	Bois, papier, imprimerie
12	Chimie, caoutchouc, plastique
13	Conseil et services informatiques, édition de logiciels
14	Conseil stratégie et organisation, prestations intellectuelles pour les entreprises
15	Construction, Architecture, urbanisme, BTP
16	Cosmétique
17	Défense et armement, police, sécurité, transport de fonds
18	Digital, e-commerce, big data, jeux électronique
19	Edition, journalisme, presse
20	Energie, eau, nucléaire, pétrole, gaz
21	Environnement, gestion des déchets
22	Equipements électriques et électroniques, composants, matériel informatique
23	Ferroviaire (matériel et équipements)
24	Formation initiale et continue, enseignement, éducation
25	Hôtellerie - restauration
26	Immobilier
27	Industries pharmaceutiques, biotechnologies, équipements médicaux
28	Industries manufacturières, mobiliers, textiles
29	Ingénierie - R&D
30	Juridique, droit et fiscalité
31	Materiel de construction
32	Matières premières, extraction, transformation, mines, métaux hors energie
33	Métallurgie et mécanique, outils
34	Mode, luxe
35	Naval
36	Négoce BtB, distribution professionnelle, import export
37	Retail, grand distribution, distribution généraliste et spécialisée
38	Santé, action sociale, hopitaux, soins, bien-être
39	Services aux entreprises (Maintenance, entretien, sécurité, travail temporaire...)
40	Services aux particuliers (cours, ménage...)
41	Sports
42	Télécoms, hébergement, internet
43	Transports marchandises, logistique, stockage, emballage, conteneurs
44	Voyages, tourisme, loisirs, jeux d'argent
45	Ressources humaines
46	Informatique
47	Support (accueil, assistanat, services généraux, ...)
48	Comptabilité, gestion, audit

Tableau 19 : tableau des secteurs d’activités

LBJ a regroupé plusieurs secteurs d'activités et les a classés en quarante-huit univers illustrés au tableau 19. Chaque univers est associé à un numéro qui nous permet de l'identifier.
Nos profils ont déjà été associés à ces univers : à chaque candidat de notre base de données, il a été associé à un secteur d'activité (univers) en tenant compte de ses récentes expériences. Nous utiliserons donc cette information afin d'interroger notre base données de telle sorte que nous ne sélectionnons que les candidats issus du secteur d'activité de notre offre d'emploi.
Notre offre d'emploi détaillée dans le tableau 18 appartient au secteur d'activité « vente » qui correspond à l'univers '37'.
Nous allons donc refaire notre processus d'appariement pour cette offre d'emploi mais en tenant compte cette fois ci du secteur d'activité.
La requête construite semi automatiquement (les compétences et le métier automatiquement et le secteur manuellement) est :

```
(expfunc:"territory manager" OR claim:"territory manager" OR pitch:"territory manager") AND skillname:(communication sales) AND univid:37
```

Le résultat de la requête comme on s'y attend ne donne que des candidats issus de l'univers 37.
Nous allons maintenant voir la différence entre ces candidats et ceux sélectionnés lors du premier processus d'appariement qui ne tenaient pas compte du secteur d'activité.

Évaluation

Même si cela est évident en observant notre requête, il faut quand même relever que les dix candidats sélectionnés sont tous issus du secteur d'activité de l'offre d'emploi.

En comparant nos dix candidats sélectionnés on remarque que :

- Ils sont tous pertinents par rapport à l'offre d'emploi : ils ont tous occupé le poste par le passé, ont tous les compétences exigées et appartiennent tous au secteur d'activité
- Ils sont différents des candidats sélectionnés par notre outil en ne tenant pas compte du secteur d'activité.

Le fait que ces dix candidats soient différents des dix meilleurs candidats sélectionnés sans tenir compte du secteur d'activité montre que les recommandations de messieurs Gravet et Tondo amélioreraient notre outil.

Nous pourrons donc envisager l'extraction automatique du secteur d'activité. L'une des méthodes serait d'étudier le vocabulaire propre à chaque univers ce qui nous permettrait de classer une offre d'emploi dans plusieurs univers avec des probabilités différentes.

VI. Conclusion

La découverte d'un nouveau domaine qui est le traitement automatique de la langue naturelle mais aussi du milieu de la recherche scientifique était un vrai défi pour la réussite de ce projet de mémoire. Nous avons appris plusieurs technologies notamment JSON, python et certaines de ses librairies, les techniques du TALN etc... Ce qui est, sur le plan personnel, une vraie valeur ajoutée pour la suite de notre carrière professionnelle.

En ce qui concerne l'appariement des offres d'emploi, notre travail est à ce jour le seul qui mesure les tâches d'extraction automatique des informations telles le métier, les compétences et l'expérience dans une offre d'emploi ce qui ne nous permet pas de comparer nos résultats avec ceux d'autres auteurs. Cependant, les scores obtenus par nos extracteurs, nous permettent de les utiliser pour notre tâche d'appariement.

La validation par des experts en recrutement de notre modèle et des résultats d'appariement est une grande satisfaction et une bonne source de motivation pour continuer à améliorer ce système.

Notre plus grande difficulté était dans la manipulation de données et de ressources avec plein de bruit; nous estimons aussi qu'avec plus de données propres, nous aurions de meilleurs résultats. Nous avons réussi aussi à constituer des ressources qui peuvent être utilisées par la communauté scientifique.

Certaines améliorations pourraient perfectionner notre outil d'appariement notamment la prise l'extraction automatique du secteur d'activité, l'utilisation de corpus parallèles pour la traduction du métier et des compétences en vue d'étendre nos requêtes mais aussi l'extraction de compétences qui ne figuraient pas au départ dans notre liste de compétences. Pour cette dernière, nous pourrons utiliser des techniques d'apprentissage machine. D'autres intervenants du projet BPP ont expérimenté pour d'autres buts et on pourrait s'inspirer de leurs études.

La méthode scientifique étant validée par ce mémoire, nous pourrons voir avec LBJ de son déploiement possible en tenant cette fois compte des contraintes technologiques avec plus d'ingénierie logicielle.

Bibliographie

Adomavicius, G. and A. Tuzhilin (2005). "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions." Knowledge and Data Engineering, IEEE Transactions on 17(6): 734-749.

Baeza-Yates, R. and B. Ribeiro-Neto (1999). "Retrieval evaluation." Modern information retrieval: 73-97.

De Campos, L. M., J. M. Fernández-Luna, et al. (2010). "Combining content-based and collaborative recommendations: A hybrid approach based on Bayesian networks." International Journal of Approximate Reasoning 51(7): 785-799.

Diaby, M. and E. Viennet "Développement d'une application de recommandation d'offres d'emploi aux utilisateurs de Facebook et LinkedIn." EGC 2014.

le Vrang, M., A. Papantoniou, et al. (2014). "ESCO: Boosting Job Matching in Europe with Semantic Interoperability." Computer(10): 57-64.

Séguéla, J. (2012). Fouille de données textuelles et systèmes de recommandation appliqués aux offres d'emploi diffusées sur le web, Paris, CNAM.