

# PERSPECTIVES DE L'IMT

## De l'autre côté du miroir Évaluer les mesures de compétences grâce aux technologies de pointe

### Principaux constats

- Les progrès en matière de puissance de calcul, de gestion de données et d'apprentissage automatique permettent une collecte et une analyse efficaces et automatisées à grande échelle des données provenant des offres d'emploi en ligne.
- Les données sur les offres d'emploi en ligne présentent deux avantages importants qui en font un complément précieux aux statistiques officielles sur les postes vacants :  
1) Elles fournissent des informations locales (p. ex., pour une municipalité) et granulaires (p. ex., pour des professions précises) en temps quasi réel sur les exigences des emplois que les employeurs posent, notamment les compétences. 2) Elles peuvent être analysées à une fraction du coût des méthodes d'enquête traditionnelles.
- Lorsqu'on utilise des données tirées d'offres d'emploi en ligne, il faut tenir compte de certaines limites importantes. La première a trait à leur représentativité, car les données d'offres d'emploi en ligne peuvent être biaisées en faveur de professions de certains secteurs et régions ainsi que selon la taille d'entreprise et les exigences de scolarité. Ensuite, certaines entreprises (p. ex., les PME ou celles de secteurs comme l'agriculture) sont plus susceptibles d'embaucher par bouche-à-oreille que d'afficher leurs postes en ligne. De plus, les différences de méthodes de traitement de données entre les fournisseurs peuvent faire varier les résultats quant aux exigences professionnelles et en matière de compétences. Enfin, il n'y a aucun moyen de savoir quelles exigences sont essentielles pour le poste en question. En fait, la façon dont une offre d'emploi est écrite ne reflète pas nécessairement le travail réellement requis. Les données permettent seulement de constater que certaines exigences sont plus fréquemment énoncées par les employeurs dans les offres d'emploi en ligne.

- Les données obtenues à partir des offres d'emploi en ligne offrent une occasion exceptionnelle de compléter les sources de données existantes sur les exigences des emplois en matière de compétences. Comme pour toutes les sources de données, y compris O\*NET, il convient de faire preuve de transparence quant aux mises en garde et aux limites afin d'éclairer la prise de décisions.

## Introduction

Au cours de la dernière année, le CIMT a travaillé sur un **projet conjoint** avec **Emploi et Développement social Canada** (EDSC) et **Statistique Canada** pour décrire les emplois en fonction des compétences requises ainsi que d'autres caractéristiques des emplois ou des travailleurs. Une approche possible **précédemment explorée** consiste à exploiter l'étendue des données disponibles grâce à l'Occupational Information Network (O\*NET) des États-Unis. Le présent *Rapport de perspectives de l'IMT* se concentre sur une autre approche : la collecte, l'analyse et la structuration des exigences professionnelles, dont les compétences, provenant des offres d'emploi en ligne, au moyen d'approches de pointe comme l'apprentissage automatique, l'accès aux Interface de programmation (API) et le moissonnage du web (voir l'**encadré 1**).

## Des explications, s'il vous plaît

La popularisation des langages de programmation libres, les progrès en matière de puissance de calcul et les améliorations de la gestion de l'information ont permis l'acquisition efficace et automatisée à grande échelle de données

### Encadré 1 : Quelles sont les technologies de pointe au 21<sup>e</sup> siècle?

#### Apprentissage automatique

L'apprentissage automatique (apprentissage artificiel, apprentissage machine) est un ensemble d'approches modernes de l'analyse statistique qui s'appuie sur des algorithmes pour traiter de grands ensembles de données. Différent des autres approches statistiques, l'apprentissage automatique repose sur des algorithmes conçus pour devenir plus précis dans la prévision des résultats (par exemple la classification de texte) sans être explicitement programmés pour le faire. En ce sens, l'apprentissage automatique est un sous-secteur de l'intelligence artificielle (IA). Dans le cas des offres d'emploi en ligne, les algorithmes sont faits pour associer les offres d'emploi à une profession et pour classer le texte brut selon un ensemble d'exigences de travail (voir l'encadré 2).

#### Moissonnage du web

Selon Statistique Canada, le moissonnage du web est un processus par lequel des renseignements sont recueillis sur le web aux fins d'extraction et d'analyse. L'une des applications du moissonnage du web consiste à collecter des données sur des sites d'emplois et des sites web d'entreprises, puis à « nettoyer » le texte. Ce processus peut être effectué à l'interne ou à l'externe. En ce moment, plusieurs entreprises d'analyse de données (par exemple : Vicinity Jobs, Burning Glass Technologies et TalentNeuron) collectent et analysent les données des offres d'emploi provenant de nombreux sites web d'entreprises et agrégateurs canadiens (comme indeed.ca).

#### Interfaces de programmation d'application (API)

Les API sont des logiciels intermédiaires qui permettent à un programme informatique externe d'obtenir de l'information d'une source interne par le biais d'un ensemble de protocoles définissant le type de données auxquelles on peut accéder. Les API permettent à des systèmes (ou des langages informatiques) indépendants de se parler. Par exemple, peu importe si vous utilisez un ordinateur Windows ou un iPhone pour accéder aux informations sur les vols, les données peuvent vous être fournies par le même point d'accès d'API. De nombreux agrégateurs d'offres d'emplois rendent les API accessibles pour que les tiers puissent plus facilement accéder à des informations structurées et les télécharger.

provenant de sources en ligne. Parallèlement, les avancées dans le domaine de l'apprentissage automatique nous ont permis d'analyser ces grands ensembles de données pour générer de nouvelles connaissances en temps quasi réel.

De plus en plus, les données brutes des offres d'emploi en ligne sont exploitées grâce à l'apprentissage automatique afin de compléter les statistiques officielles destinées à l'analyse du marché du travail. Puisque ces données, disponibles en temps réel et selon le lieu précis (par exemple la ville), sont recueillies à partir de sites web accessibles au public, elles contribuent à améliorer notre accès à une information sur le marché du travail (IMT) plus locale, granulaire et actuelle. De plus, elles offrent l'occasion d'explorer les données sur les compétences du point de vue de la recherche de talents par les employeurs.

### **Des exigences professionnelles aux informations sur les compétences**

Lorsque des données non structurées sont collectées à partir d'une offre d'emploi en ligne, le texte doit être analysé et classé par catégories. Une technique répandue d'analyse de texte est le **Document Object Model** (DOM), qui permet à un programme d'extraire de l'information d'une page web en référant son emplacement dans le document. Cette méthode, cependant, renvoie toutes les données à l'endroit spécifié. Autrement dit, le résultat consiste simplement en du texte brut qui doit encore être formaté pour l'analyse. Pour ce faire, les entreprises d'analyse de données appliquent généralement des algorithmes de traitement du langage naturel (TLN) poussés et souvent propriétaires qui utilisent l'apprentissage automatique pour classer le texte non structuré selon une taxonomie des compétences et des exigences de travail préexistante (voir l'**encadré 2**).

### **Encadré 2 : Que sont les taxonomies et pourquoi sont-elles utiles?**

Dans sa forme la plus simple, une taxonomie est un système de classification. En matière de compétences, les taxonomies servent à organiser les informations les plus diverses en catégories utilisables. Les taxonomies fournissent également un langage commun pour discuter des compétences requises pour les emplois et les professions.

Parmi les différentes taxonomies de compétences actuelles, l'une des plus populaires est intégrée dans le  système O\*NET des États-Unis. Citons aussi la  Classification européenne des aptitudes/compétences, certifications et professions  et la nouvelle  Taxonomie des compétences et des capacités.

### **De quoi ont l'air les données?**

Pour donner un aperçu des données disponibles, nous présentons un résumé de l'information provenant de 2 467 709 offres d'emplois en ligne de partout au Canada en 2019. Les résultats précis figurant ici sont ceux de l'entreprise d'analyse de mégadonnées **Vicinity Jobs**, mais le type et la fréquence d'information disponibles sont les mêmes pour tous les fournisseurs de données<sup>1</sup>. Ces données ont déjà été nettoyées et structurées par l'entreprise de moissonnage. Les principales variables d'intérêt et leur fréquence d'apparition sont présentées dans le **tableau 1**. Les informations sur les offres d'emploi de Vicinity Jobs seront disponibles sur le futur tableau de bord des offres d'emploi du CIMT.

**Tableau 1 : Disponibilité des informations contenues dans les données de Vicinity Jobs (2019) par variable**

Catégorie	Variable	Description	Pourcentage des offres d'emploi
<b>Employeur</b>	Nom de l'employeur	Nom de l'employeur responsable de l'annonce	39 %
<b>Lieu</b>	Ville ou municipalité	Ville ou localité où l'emploi est situé	91 %
	Division du recensement	Division de recensement où l'emploi est situé	93 %
	Région économique	Région économique où l'emploi est situé	93 %
	Province ou territoire	Province où l'emploi est situé	100 %
<b>Secteur d'activité (SCIAN)</b>	Code SCIAN à 6 chiffres	Classe précise du secteur d'activité de l'offre d'emploi	43 %
<b>Profession (CPN)</b>	Code CNP à 1 chiffre	Classification professionnelle générale de l'offre d'emploi	87 %
	Code CNP à 4 chiffres	Classification professionnelle précise de l'offre d'emploi	71 %
<b>Expérience</b>	Expérience	Expérience explicitement mentionnée comme une exigence	16 %
<b>Éducation</b>	Éducation	Type de formation requis pour le poste (p. ex. : diplôme d'études secondaires, diplôme d'études supérieures, etc.)	42 %
<b>Certification</b>	Certification	Licence, certification, développement professionnel requis	23 %
<b>Salaire offert</b>	Salaire	Rémunération mentionnée dans l'offre d'emploi	16 %
<b>Durée</b>	Emploi permanent ou temporaire		38 %
<b>Type</b>	Emploi à temps plein ou à temps partiel		95 %
<b>Exigences professionnelles</b>	Outils, compétences, connaissances, technologies et autres descripteurs nommés par l'employeur comme exigences de l'emploi	L'ensemble des exigences de travail* classés selon la Taxonomie des compétences et des capacités d'EDSC : 1) Connaissances 2) Compétences 3) Outils et technologies 4) Autres <sup>2</sup>	90 %

Source : Vicinity Jobs (2019).

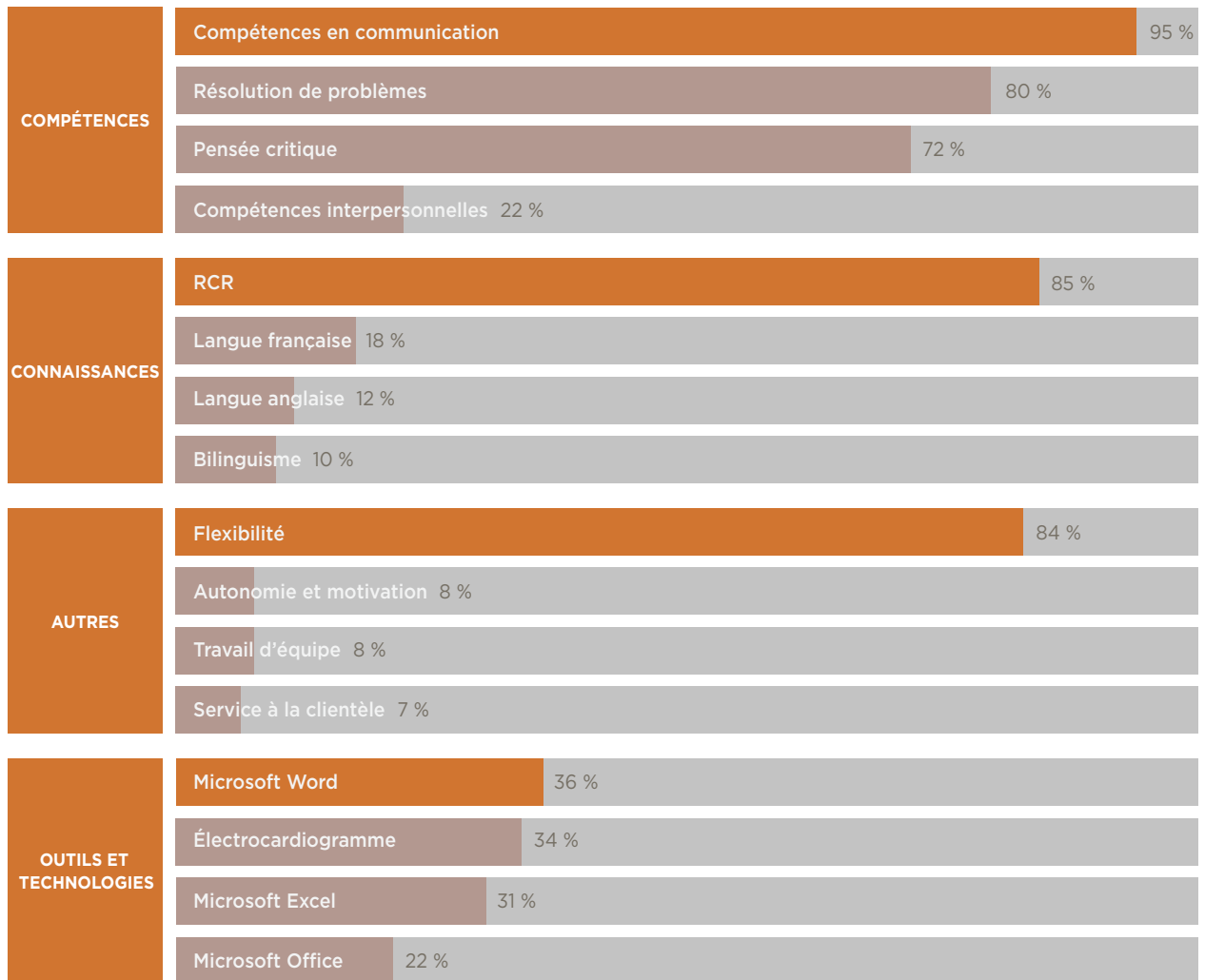
\* Structurées par le CIMT en partenariat avec Vicinity Jobs et EDSC.

## Quels renseignements peut-on tirer de ces données?

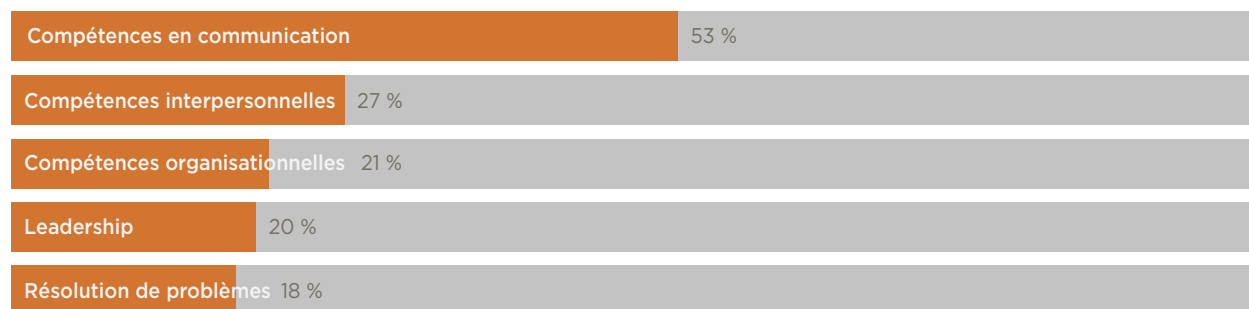
Comme l'indique le **Rapport de perspectives de l'IMT n° 14**, l'un des principaux avantages de recueillir et d'analyser des données provenant d'offres d'emplois en ligne est la possibilité de connaître les exigences de travail que les employeurs nomment explicitement. À l'aide des données fournies par Vicinity Jobs, on peut calculer la part des exigences de travail spécifiques (par exemple : connaissances, compétences, outils et technologies) énumérées par les employeurs pour différents lieux et professions. La **figure 1** présente les quatre principales exigences de travail

par groupe (c'est-à-dire : compétences, connaissances, outils et technologies, autres) pour les infirmiers/infirmières (CNP 3012) à Winnipeg. On peut également examiner une catégorie en particulier, comme les compétences. La **figure 2** présente les cinq principales compétences parmi toutes les offres d'emplois à Winnipeg en 2019. De même, à la **figure 3**, on peut voir les cinq principaux outils et technologies exigés pour les programmeurs/programmeuses et développeurs/développeuses en médias interactifs (CNP 2174).

**Figure 1 : La communication, le RCR, la flexibilité et Microsoft Word sont en tête des compétences requises pour les infirmiers autorisés/infirmières autorisées et infirmiers psychiatriques autorisés/infirmières psychiatriques autorisées (CNP 3012) à Winnipeg (2019).**



**Figure 2 : Compétences en communication sont les compétences les plus fréquemment recherchées à Winnipeg parmi toutes les offres d'emploi (2019)**



**Figure 3 : Le logiciel Git est l'outil technologique le plus fréquemment recherché pour les programmeurs/programmeuses et développeurs/développeuses en médias interactifs (CNP 2174) à Winnipeg (2019).**



## Avantages et limites

Les données recueillies à partir des offres d'emploi en ligne offrent plusieurs avantages importants, le plus évident étant la possibilité de se faire une idée en temps quasi réel des exigences professionnelles des employeurs. Cela permet de cerner les connaissances, les compétences, les outils, les technologies, etc. dont les chercheurs d'emploi ont besoin pour réussir dans le marché du travail d'aujourd'hui. Comme ces informations proviennent des offres affichées sur des sites d'emplois publics, il n'existe aucune restriction qui limiterait le niveau de détail accessible.

Les offres d'emploi en ligne sont également utiles en raison du grand volume de données et de l'information localisée qu'elles fournissent. Par exemple, une analyse des données 2019 de

Vicinity Jobs, montre que 91 % des offres d'emploi précisent la ville ou la municipalité, ce qui en fait la source de données la plus locale d'informations sur les exigences professionnelles actuellement disponibles. De plus, on peut obtenir ce volume de données à une fraction du coût des méthodes traditionnelles de sondage. Les algorithmes d'apprentissage automatique utilisés pour parcourir les pages web, par exemple, peuvent être implantés avec des langages de programmation libres (disponibles gratuitement et publiquement). Les millions de points de données disponibles permettent potentiellement d'obtenir une plus grande fiabilité statistique et de regrouper les observations selon les besoins pour une analyse personnalisée des données.

Ces informations sont précieuses pour un grand nombre d'intervenants, notamment les éducateurs qui mettent au point des programmes de cours, les étudiants qui choisissent leur parcours scolaire et de formation, les décideurs politiques qui allouent des fonds pour l'amélioration des compétences et les organisations qui travaillent à la planification de la main-d'œuvre. En outre, l'examen de l'évolution des exigences de travail au fil du temps permet de repérer rapidement les tendances émergentes, comme la récente augmentation de la demande de connaissances en intelligence artificielle pour les actuels.

L'utilisation des données relatives aux offres d'emploi en ligne comporte aussi d'importantes limites, notamment en ce qui concerne leur représentativité. Des études antérieures comparant les données sur les offres d'emploi à celles tirées d'enquêtes nationales sur le marché du travail ont mis en évidence plusieurs différences. Les grands employeurs, **par exemple**, sont plus susceptibles de publier toutes les offres d'emploi en ligne, tandis que les plus petites entreprises n'y affichent que les postes de direction. De même, les professions dans des domaines comme les soins de santé et les technologies de l'information sont plus susceptibles d'être surreprésentées, tandis que celles dans des domaines comme la construction sont sous-représentées. Les postes de direction et ceux qui nécessitent généralement des études universitaires sont presque deux fois plus susceptibles d'être affichés en ligne sur Vicinity Jobs qu'ils ne figurent dans l'**Enquête sur les postes vacants et les salaires** (EPVS), un sondage national mené tous les trimestres par Statistique Canada.

Selon les secteurs, les employeurs recrutent de différentes manières. Les petites entreprises et celles qui œuvrent dans des domaines comme l'agriculture sont plus susceptibles d'embaucher par le bouche-à-oreille que d'annoncer leurs postes en ligne. Pour compliquer les choses encore

davantage, les sources en ligne comme LinkedIn utilisent des technologies antimoissonnage afin d'empêcher les tiers d'extraire leurs données à grande échelle. De plus, plusieurs entreprises d'analyse de données omettent les annonces en français. Toutes ces limitations peuvent rendre difficiles l'analyse des tendances et les comparaisons dans le temps, car la taille de l'échantillon et la composition des affichages collectés peuvent varier d'une année à l'autre.

Ensuite, les méthodes de traitement du langage naturel qui structurent le texte brut des annonces varient selon les fournisseurs. Comme beaucoup de ces méthodes sont propriétaires, elles ne sont pas soumises à l'examen du public. L'utilisation de méthodologies ouvertes et transparentes est une étape importante pour créer des données fiables et de qualité, mais cela reste un défi.

De plus, il n'y a aucun moyen de savoir quelles exigences sont essentielles pour le poste en question. Les données permettent seulement de constater que certaines exigences sont plus fréquemment énoncées par les employeurs dans les offres d'emploi en ligne.

Enfin, les offres d'emploi en ligne peuvent être biaisées de différentes manières, par exemple par l'omission d'informations ou l'inflation des compétences. L'omission d'informations est particulièrement probable si les employeurs s'attendent à ce que certaines exigences soient évidentes pour les candidats. Par exemple, il va de soi que la numérotation est essentielle pour les ingénieurs, mais elle est rarement mentionnée dans l'offre d'emploi. Par ailleurs, certains employeurs peuvent préciser des exigences qui dépassent les besoins réels du poste. De tels biais fausseront les résultats des analyses des exigences professionnelles si seules les données des offres d'emploi en ligne sont utilisées.

## La voie à suivre

Étant donné leur actualité, leur granularité, leur caractère local et leur fréquence, les données tirées des offres d'emploi en ligne offrent une précieuse occasion de compléter les sources de données existantes. À preuve, ce type de données est largement utilisé par les intervenants comme les agences gouvernementales, les conseils de gestion de la main-d'œuvre et les entreprises privées pour améliorer entre autres la planification de carrière, la formation axée sur les compétences et le développement de programmes. Aussi, les données offrent un aperçu des compétences précises recherchées par les employeurs ainsi que des autres exigences professionnelles qu'on retrouve sur le marché du travail.

Cependant, comme pour toutes les sources de données, des stratégies améliorées sont nécessaires pour remédier aux limites comme le manque de représentativité et la sous-estimation ou la surestimation des exigences en matière de compétences. Une solution que le CIMT, EDSC et Statistique Canada exploreront dans un prochain *Rapport de perspectives de l'IMT* les avantages et les limites d'interroger directement les employeurs sur leurs exigences en matière de compétences, comme on le fait maintenant au Royaume-Uni et en Australie.

## Remerciements

Ce *Rapport de perspectives de l'IMT* a été préparé conjointement par le personnel du Conseil de l'information sur le marché du travail, Emploi et développement social Canada (Direction de l'information sur le marché du travail) et Statistique Canada (Centre de l'information sur le marché du travail). Nous tenons à remercier David Ticoll (University of Toronto), Jacob Loree (**Université Ryerson**), ainsi que Ron Samson et Austin Hracs (**Magnet**) pour leurs commentaires.

Vos commentaires sont les bienvenus. Nous vous invitons à nous faire part de vos observations et opinions sur le sujet en nous écrivant à [info@lmic-cimt.ca](mailto:info@lmic-cimt.ca).

## Notes

- 1 Vicinity Jobs, l'un des principaux fournisseurs de données sur les offres d'emplois en ligne, a établi un partenariat avec le CIMT pour exploiter cette source de données à des fins d'analyse du marché du travail. Bien que notre analyse repose principalement sur les données de Vicinity Jobs, les limites mentionnées dans le présent rapport sont les mêmes pour tous les fournisseurs de données (**Burning Glass Technologies**, **Talent Neuron**, etc.).
- 2 Vicinity Jobs relie d'abord les offres d'emploi aux exigences de travail au moyen de sa propre taxonomie propriétaire, classifiant ainsi les données texte brutes en cinq groupes : 1) compétences générales/non techniques, 2) compétences spécialisées, 3) technologies, 4) outils et équipement et 5) autres. En partenariat avec Vicinity et EDSC, le CIMT a reclassifié chaque exigence de travail selon la Taxonomie des compétences et des capacités d'EDSC : 1) connaissances, 2) compétences, 3) outils et technologies et 4) autres.