

Connaissance de l'emploi

Le 4 pages du Centre d'études de l'emploi et du travail
Mai 2017

136 le **cnam**
ceet

GOOGLE ET LE MARCHÉ NUMÉRIQUE DU TRAVAIL

Yannick Fondeur,
Cnam, Lise/CNRS UMR 3320, CEET

Principale porte d'entrée sur le *web*, *Google* a indirectement contribué à structurer le marché du travail en ligne par l'intermédiaire de ses algorithmes de classement et de son programme d'achat de mots-clés, qui déterminent le référencement des sites d'emploi. Plusieurs initiatives récentes indiquent que le géant du *web* souhaite maintenant intervenir *directement* sur le champ de l'emploi. C'est sur l'une d'entre elles que se focalise ce 4-pages.

Dévoilée fin 2016, la *Cloud Jobs API* propose aux recruteurs et aux candidats un « langage commun » facilitant leur coordination *via* les services de diffusion d'offres d'emploi et de CV. Une large adoption de ce dispositif aurait des conséquences importantes sur l'activité des acteurs du marché numérique du travail. L'initiative de *Google* conduit également, en creux, à s'interroger sur le rôle des services publics de l'emploi dans la régulation et la mise en forme du marché du travail en ligne.

Google est la principale porte d'entrée sur le *web*. Les moteurs de recherche constituent en effet la première source de trafic pour les sites *web*, et l'entreprise californienne détient en France près de 95 % de ce marché. Être visible sur ses premières pages est donc un enjeu majeur. Il y a deux manières de l'être : le référencement payant qui consiste à sponsoriser l'information *via* le système *Adwords*¹ et le référencement dit « naturel ». Ce dernier n'a en fait rien de « naturel » : d'une part, parce qu'il est le fruit des choix et représentations incorporés dans les algorithmes d'indexation et de classement ; d'autre part, parce que ces artefacts cognitifs engendrent des comportements d'adaptation des éditeurs de sites. Avec son système de classement des pages *web*, *Google* a à la fois « imposé au *web* un ordre qui lui est propre » et engendré un perpétuel « jeu du chat et de la souris »

avec les stratégies de référencement, qui conduit ses ingénieurs à le modifier sans cesse (Cardon, 2013).

Google est également, pour les candidats à l'embauche, un mode d'entrée sur le marché numérique du travail. Il est ainsi significatif que l'analyse des requêtes lancées à partir du mot-clé « emploi » permette de retracer avec précision et en temps réel l'évolution du chômage des jeunes (Fondeur & Karamé, 2013) : une hausse du chômage entraîne, pour cette population qui mobilise beaucoup Internet, un accroissement immédiat des recherches contenant ce mot-clé. Pour autant, *Google* ne propose pas, pour le moment, d'outil de recherche dédié à l'emploi, laissant le champ libre aux agrégateurs d'offres d'emploi. Toutefois, c'est l'objet de ce *Connaissance de l'emploi*, une initiative dévoilée fin 2016 suggère que la situation pourrait évoluer.

¹ Les termes techniques liés au *web* sont explicités dans le glossaire.

GLOSSAIRE

SEO (*Search Engine Optimization*) : ensemble de techniques visant à optimiser le référencement « naturel » d'un site, en agissant sur son contenu, sa structure et ses liens avec le reste du *web*, en fonction des critères supposés déterminants dans les algorithmes des moteurs de recherche.

Adwords : dispositif permettant de sélectionner des termes de recherche sur *Google* et d'y associer une annonce qui apparaîtra dans les résultats lorsque les mots-clés en question seront utilisés par les internautes. L'annonceur alloue un budget à sa campagne, qui n'est débité que lorsque des internautes cliquent effectivement sur ses liens. Le « coût par clic » (**CPC**) dépend de différents facteurs, dont l'intensité de la demande sur les termes de recherche sélectionnés.

SEM (*Search Engine Marketing*) : ensemble des actions mises en œuvre pour améliorer son référencement sur les moteurs de recherche (SEO et achat des mots-clés *via* des dispositifs type *Adwords*).

XML (*Extensible Markup Language*) : métalangage informatique générique, mis au point sous l'égide du *World Wide Web Consortium* (W3C), facilitant l'échange de contenus complexes entre des systèmes d'information hétérogènes.

API (*Application Programming Interface*) : dispositif permettant la communication et l'échange de données entre systèmes tiers par le biais d'un service *web*.

dans les premières positions. Les candidats à l'embauche sont ainsi davantage enclins à cliquer sur ses liens que sur ceux des acteurs traditionnels. On touche là au cœur du modèle d'activité de l'agrégateur qui, de fait, est parvenu à imposer un niveau d'intermédiation supplémentaire entre *Google* et les diffuseurs d'offres d'emploi.

Pour y parvenir, *Indeed* consacre certes un budget conséquent à l'achat sur *Adwords* de mots-clés liés à l'emploi, mais **l'agrégateur est également maître dans l'art de s'adapter aux algorithmes de classement de *Google*** pour optimiser son référencement « naturel » (SEO). La clé de son succès serait ainsi une connaissance intime des critères de classement du moteur de recherche. « Il y a beaucoup d'anciens de *Google* chez *Indeed*, donc je pense qu'ils connaissent parfaitement la manière dont l'algorithme est pensé », confie un responsable de site d'emploi classique. Il en veut pour preuve le fait que l'agrégateur se serait parfaitement adapté à *Panda*, un algorithme introduit à partir de 2011 par *Google* pour filtrer les pages ne proposant pas de contenu textuel suffisant et/ou dupliquant des textes publiés ailleurs, des critères *a priori* défavorables aux agrégateurs : « tous les agrégateurs, après *Panda*, se sont pris des baisses de trafic, très fortes (...) ; et le seul, bizarrement, qui a réussi à contourner un peu cette nouvelle règle de *Panda*, c'est *Indeed* ».

***Indeed.fr* est, depuis fin 2014, le second site d'emploi français le plus visité**, derrière Pôle emploi et devant *Leboncoin*. Cette montée en puissance rapide s'est faite pratiquement sans dépenses de communication, à l'inverse des sites d'emploi classiques, qui y consacrent d'importants budgets. *Indeed* n'a lancé sa première campagne de communication en France que fin 2015 et n'y dispose d'un service de presse que depuis début 2016. La qualité de son référencement sur *Google* lui a permis d'attirer une audience supérieure aux grands sites d'emploi classiques, malgré une faible notoriété spontanée. Son modèle économique s'appuyant sur la vente de clics, on peut considérer que c'est ce référencement qu'il revend aux diffuseurs classiques d'offres d'emploi. Ce faisant, une partie des revenus que *Google* pourrait tirer de ces derniers *via Adwords* est en quelque sorte captée par l'agrégateur. Pour l'illustrer, prenons l'exemple d'un des grands sites d'emploi français, *Regionsjob*, qui a rendu publiques ses principales sources de trafic pour avril 2016 : à cette date, 22 % de son audience provenait du référencement « naturel » de ses pages sur les moteurs de recherche (essentiellement *Google*), 13 % des achats de mots-clés sur *Adwords* et 20 % des agrégateurs, dont 12 % pour le seul *Indeed*².

Encadré 2

L'ENQUÊTE

L'enquête s'appuie sur une veille des dispositifs numériques d'intermédiation depuis 2004 et sur environ quatre-vingts entretiens, informels ou semi-directifs, menés depuis cette date auprès d'acteurs de l'emploi en ligne dans le cadre de différents projets de recherche. Les *verbatim* proviennent tous d'une recherche menée en 2015-2016 qui a bénéficié d'une subvention de Pôle emploi (Fondeur, 2016a). Ce travail a fait l'objet d'un précédent *Connaissance de l'emploi* (Fondeur, 2016b), centré sur la « transparence » du marché numérique du travail, qui constitue une lecture complémentaire à ce numéro.

● Les agrégateurs d'offres d'emploi, « *Google* de l'emploi » ?

Nés au début des années 2000, les **agrégateurs d'offres d'emploi indexent les annonces publiées sur le *web*** et offrent aux candidats la possibilité d'effectuer leurs recherches depuis un point unique. Se présentant volontiers comme des « *Google* de l'emploi », ils reprennent généralement plusieurs caractéristiques du moteur de recherche généraliste : une page d'accueil dotée d'une interface très simple invitant à entrer des mots-clés en texte libre ; des pages de résultats n'affichant que les premières lignes des offres indexées, suivies de liens vers les sites émetteurs ; un système de *sponsoring* des offres au coût par clic (CPC).

Les **agrégateurs se positionnent donc en tant qu'intermédiaires informationnels** entre les diffuseurs d'offres d'emplois et les candidats, même s'il s'agit, pour une large part, d'une intermédiation indirecte, passant par *Google*. Ils y sont en effet généralement très bien référencés et captent ainsi une proportion importante des clics des internautes recherchant un emploi *via* le moteur de recherche généraliste.

C'est en particulier le cas d'*Indeed*, leader mondial du secteur : une grande partie des requêtes formulées sur *Google.fr* contenant le mot-clé « emploi », un terme professionnel et éventuellement un lieu (par exemple : « emploi commercial Paris ») place ses pages

● Les offres d'emploi sur Internet : dispersion et hétérogénéité

En dehors de la qualité de leur référencement sur les moteurs de recherche généralistes, le succès des agrégateurs s'explique par la topographie particulière du marché numérique du travail. En effet, loin de faire advenir une bourse à l'emploi globale, **le développement d'Internet a favorisé une diffusion décentralisée des offres** qui se trouvent dispersées sur une multitude de supports, configuration qui rend pertinent le service de « recentralisation *a posteriori* » des offres proposé par les agrégateurs. Plusieurs services publics de l'emploi ont d'ailleurs considéré qu'une telle activité relevait de leurs missions. En Europe, cela a été le cas de la *Bundesagentur für Arbeit* en Allemagne, puis, plus récemment, de Pôle emploi en France. Depuis 2013, ce dernier met en œuvre un dispositif d'agrégation dans le cadre de sa démarche de « Transparence du Marché du Travail » (Fondeur, 2016a, 2016b).

Si la dispersion des supports de diffusion des offres justifie l'existence de services d'agrégation, privés ou publics, une autre caractéristique du marché du travail en ligne rend, elle, l'activité des agrégateurs complexe et incertaine : **l'absence de langage commun pour décrire les offres**. Cette absence est due, d'une

² Le reste provenait essentiellement des clics sur les liens des alertes *email* et de l'accès direct.

part, à une hétérogénéité syntaxique : malgré quelques tentatives de standardisation, les offres d'emploi sont publiées selon divers *formats* de données, qui nuisent à leur circulation d'une base à l'autre. Elle relève, d'autre part, d'une hétérogénéité de catégorisation : il n'existe pas de *lexique* partagé par les acteurs pour qualifier les emplois, notamment en termes de métiers et de compétences. La structure d'une offre d'emploi étant assez simple et faisant l'objet de conventions, l'absence de *format* standard ne pose pas d'incommensurables problèmes aux systèmes modernes d'extraction de l'information, d'autant que le langage XML (*cf.* glossaire), qui s'est généralisé, facilite les échanges de données.

Par contre, la question de la catégorisation est plus ardue. Tous les acteurs n'y apportent d'ailleurs pas le même type de réponse. En effet, **le mécanisme mis en place par Pôle emploi depuis 2013 diffère fortement de celui des agrégateurs privés**. Alors que ces derniers évacuent le problème en s'appuyant sur le libellé des offres et proposent une recherche de mots-clés en plein texte, l'opérateur public tente d'imposer sa catégorisation des offres, notamment sa nomenclature des métiers et des emplois, le Rome³. Cette façon de faire présente l'avantage de fournir aux candidats un système de classement des offres leur permettant de circonscrire précisément leurs recherches, mais elle suppose de passer par un algorithme sémantique capable de lire l'offre et de l'interpréter pour la qualifier dans la « langue » de Pôle emploi, c'est-à-dire pour lui affecter un code Rome. Or, c'est une opération très destructrice qui aboutit à la perte d'un nombre important d'offres d'emploi pour lesquelles le système ne parvient pas à identifier un code Rome et qui ne peuvent donc être agrégées. Par ailleurs, certaines offres effectivement agrégées se retrouvent mal catégorisées.

● La *lingua franca* du marché numérique du travail ?

C'est précisément cette question du langage commun qui est à l'origine d'une initiative rendue publique par *Google* en fin d'année dernière et qui sous-tend le premier pas du moteur de recherche généraliste dans le domaine de l'emploi. L'ambition affichée n'est pas moins que de « *fournir la nécessaire lingua franca entre la personne à la recherche d'un emploi et l'offre d'emploi de l'employeur* »⁴. L'analogie est intéressante. La langue franque est un *pidgin*, c'est-à-dire une langue véhiculaire simplifiée, qui a été utilisée du Moyen Âge au XIX^e siècle par les marins et marchands circulant dans le bassin méditerranéen. Le *Dictionnaire universel* d'Antoine Furetière (1690) le décrit comme « *un jargon qu'on parle sur la mer Méditerranée, composé de français, d'italien, d'espagnol et d'autres langues, qui s'entend par tous les matelots et marchands de quelque nation qu'ils soient* ».

Baptisé *Cloud Jobs API*, le dispositif de *Google* fait actuellement l'objet d'un test fermé auprès de grandes entreprises et d'acteurs du marché numérique du travail. À ce stade, la firme californienne le présente comme destiné aux acteurs finaux, recruteurs et candidats, auxquels elle entend fournir un langage commun pour désigner les emplois et les compétences, dans le but d'éviter les défauts d'appariement liés à des façons différentes d'exprimer la même réalité (*semantic mismatch*). Le dispositif est interrogeable par le biais d'une API (*cf.* glossaire) à laquelle les acteurs du marché numérique du travail devraient avoir accès à l'issue du test.

Pour créer sa *lingua franca* du marché du travail, qui demeure à ce stade une langue anglaise, *Google* a utilisé deux types de sources : des nomenclatures préexistantes en libre accès (notam-

ment la classification O*Net-SOC⁵ développée par le Département du Travail des États-Unis) et un ensemble de dix-sept millions d'offres d'emploi collectées sur le *web* américain et dont ont été extraits des intitulés d'emplois et de compétences. S'appuyant sur des techniques *big data*, *Google* a construit à partir de ce matériau un **dictionnaire d'emplois** ainsi qu'un **dictionnaire de compétences** et modélisé les relations entre ces deux champs pour faire apparaître les compétences associées aux différents emplois.

Les deux dictionnaires sont **structurés comme des ontologies**. À la différence d'une nomenclature classique qui propose une classification hiérarchique fondée sur des choix raisonnés, une ontologie est une mise en forme moins « autoritaire » qui laisse la possibilité d'employer plusieurs termes pour désigner la même chose. Outil bien connu en informatique et en sciences de l'information⁶, elle permet de formaliser l'ensemble des relations sémantiques existant entre différentes catégories conceptuelles d'un champ : des relations de subsomption totale⁷, hiérarchiques, comme dans la logique ensembliste d'une nomenclature classique, mais aussi des relations de subsomption partielle, c'est-à-dire d'équivalence, de proximité sémantique, etc. Sa souplesse en fait un dispositif parfaitement adapté à la coordination (alors que la nomenclature est plus adaptée à la statistique).

Ainsi, l'ontologie de compétences élaborée par *Google* comprendrait 50 000 catégories conceptuelles reliées entre elles par six types de relations, ce qui correspond à un graphe sémantique complexe et de très grande dimension. L'ontologie d'emplois et de métiers est, semble-t-il, plus classique dans sa structure, bien que contenant un nombre encore plus important de catégories conceptuelles. Elle est formée de trois niveaux, dont les deux premiers (comptant respectivement 30 et 1 100 catégories) correspondent à ceux proposés par la nomenclature O*Net-SOC. L'apport spécifique de *Google* est le troisième niveau, composé de pas moins de 250 000 intitulés d'emploi (par comparaison, le « Rome », qui est déjà une nomenclature d'une grande finesse, n'en compte que 11 000).

Cet équipement peut être analysé comme un « investissement de forme » (au sens de Thévenot, 1986) : **il s'agit pour *Google* d'investir dans la mise en forme du marché numérique du travail afin d'offrir à ses acteurs un langage commun facilitant leur coordination**. Les sites d'emploi réalisent tous ce type d'investissement, qui est au cœur de leur activité d'intermédiation (Marchal, Mellet et Rieucou, 2005). Il en est de même des entreprises et de leurs organisations professionnelles, lorsqu'elles produisent des référentiels, et, bien sûr, des services publics de l'emploi au travers de la production de nomenclatures de référence. La particularité de l'initiative de *Google* est qu'elle relève d'un *méta*-investissement capitalisant, par le biais du *big data*, sur les investissements de forme déjà réalisés par les acteurs du secteur, soit indirectement en les absorbant *via* les offres d'emploi publiées sur Internet, soit directement en intégrant des nomenclatures complètes en libre accès. Notons qu'alors même que sa construction s'appuie sur des données publiques, le dispositif du moteur de recherche relève d'un modèle propriétaire et qu'il n'est pas envisagé pour le moment de le rendre *open source*.

● Quels enjeux ?

Il est trop tôt pour juger de l'initiative de *Google* : le système testé est en version *alpha* et il faut rappeler que l'entreprise a, par le passé, abandonné de nombreux services lancés en version publique⁸.

³ Répertoire opérationnel des métiers et des emplois.

⁴ <https://cloud.google.com/blog/big-data/2016/11/cloud-jobs-api-machine-learning-goes-to-work-on-job-search-and-discovery>

⁵ <https://www.onetonline.org>

⁶ Le terme appartient initialement au domaine de la philosophie, où il revêt un sens beaucoup plus général.

⁷ Par exemple, la catégorie d'« employé de banque » est entièrement contenue dans celle, plus générale, d'« employé ».

⁸ À titre d'exemples : *Spaces*, *Wave*, *Buzz*, *Base*, *Answers*, *SearchWiki*, *Reader*, *Notes*, etc.

Néanmoins, parce que cette initiative constitue une réponse à une problématique identifiée depuis de nombreuses années (Fondeur & Tuchsirer, 2005 ; Fondeur, 2016a, 2016b) et parce qu'elle interroge, de ce point de vue, le rôle des services publics de l'emploi, il est utile de se placer dans l'hypothèse heuristique où le service serait pérennisé et ferait l'objet d'une adoption massive. Par ailleurs, le fait que ce dernier teste en parallèle un service, baptisé *Google Hire*, offrant aux entreprises des fonctionnalités de gestion de leurs recrutements en ligne, indique une stratégie globale de pénétration du marché numérique du travail.

L'intervention directe de la firme californienne sur le marché numérique du travail est, depuis plusieurs années, un sujet de discussion parmi les acteurs du secteur, du fait des répercussions importantes qu'elle pourrait avoir sur leur activité. Cette intervention a longtemps été imaginée sous la forme d'un service d'hébergement d'offres d'emploi, hypothèse soutenue par quelques incursions dans le domaine des annonces classées, comme le service *Google Base*, aujourd'hui abandonné. Cependant, l'intérêt de *Google* est sans doute différent, comme l'explique un dirigeant de site d'emploi. « *Tout le monde dit : "Mais pourquoi Google ne crée pas un site emploi ?" (...) Quand j'additionne toutes les dépenses en SEM sur Google de tous les acteurs du réseau, je me dis que Google n'a pas intérêt (...) à lancer un job board. J'avais posé la question au patron Europe, il m'avait répondu exactement ça, il m'avait dit : "Tant qu'on gagne plus d'argent avec vous – parce que c'est de la marge pure pour eux – on n'a pas intérêt à lancer un site emploi"* ».

Si le dispositif expérimenté ne percute pas frontalement l'activité des sites d'emploi classiques, il en va tout autrement pour les acteurs qui ont fondé leur activité sur les « imperfections » du marché du travail en ligne, c'est-à-dire les multidiffuseurs et les agrégateurs d'offres d'emploi (Fondeur, 2016a, 2016b). En effet, au-delà des défauts d'appariement entre candidats et recruteurs, mis en avant dans la communication de *Google*, **l'enjeu majeur de l'initiative concerne la fluidification de la circulation des informations sur le marché numérique du travail.**

Les **multidiffuseurs** d'offres d'emploi sont des acteurs qui permettent aux recruteurs d'envoyer leurs annonces vers différents sites d'emploi depuis une interface unique. Cette activité nécessite de prendre en charge l'hétérogénéité des investissements de forme réalisés par les sites interfacés. Concrètement, les multidiffuseurs disposent d'un système *interne* traduisant les offres d'une « langue » à l'autre. Or, en proposant un langage universel partagé, *Google* s'attaque au cœur de leur activité.

Les conséquences potentielles sur l'activité des **agrégateurs** sont tout aussi importantes. Il est fort probable qu'en cas de succès de son dispositif, *Google* développe un outil de recherche spécifiquement dédié aux offres d'emploi. Ouvrir un tel service serait d'ailleurs de nature à accélérer l'adoption de sa *lingua franca*, puisque son usage deviendrait la condition d'un bon référencement « naturel ». Parallèlement, *Google* pourrait également récupérer

une grande partie des dépenses d'acquisition de trafic captées par les agrégateurs.

Au-delà de l'impact sur l'ordre écologique des acteurs privés, **c'est le rôle de l'action publique dans la régulation et la mise en forme des marchés numériques du travail qui est interrogé.** Un outil de coordination, accessible à tous et facilitant la circulation de l'information, peut en effet être considéré comme un *bien public*. En France, le « Rome » de Pôle emploi pourrait remplir ce rôle. Disponible depuis 2015 en *open data*, il n'a pas fait l'objet d'une adoption massive, mais il est désormais possible pour *Google*, ou tout autre acteur, de l'intégrer à un méta-investissement de forme destiné à élaborer une *lingua franca* commune aux marchés du travail francophones. Pour autant, s'il s'agit d'un équipement d'intérêt général, la construction d'un tel outil et sa mise à disposition des acteurs pourrait relever des missions du service public de l'emploi.

Le fait qu'en France Pôle emploi se soit engagé dans une démarche de « transparence du marché du travail », fondée sur la circulation des offres et demandes d'emploi (Fondeur, 2016a, 2016b), fait d'une telle initiative un aboutissement attendu. Reste que cela supposerait d'élargir l'investissement de forme réalisé au travers du « Rome », pour lui donner une structure ontologique et en faire un dispositif plus universel. Cela reviendrait en pratique à l'alimenter par l'analyse d'offres d'emploi⁹ et l'intégration d'autres référentiels, ou, de manière plus originale, à offrir aux acteurs la possibilité d'y contribuer directement *via* l'ouverture régulée de droits en écriture.

BIBLIOGRAPHIE

- Cardon D., 2013, « Dans l'esprit du PageRank. Une enquête sur l'algorithme de Google », *Réseaux*, n° 177.
- Fondeur Y., 2016a, « Dynamiques écologiques du marché du travail en ligne autour de la circulation des offres d'emploi », in *Diversité et dynamiques des intermédiaires du marché du travail*, Pôle emploi, série « Études et Recherches ».
- Fondeur Y., 2016b, « Les offres d'emploi sur Internet : vers la "transparence" du marché du travail ? », Cnam-CEET, *Connaissance de l'emploi*, n° 132.
- Fondeur Y., Karamé F., 2013, « Can Google data help predict French youth unemployment? », *Economic Modelling*, vol. 30.
- Fondeur Y., Tuchsirer C., 2005, *Internet et les intermédiaires du marché du travail*, Rapport Ires pour l'ANPE.
- Marchal E., Mellet K. et Rieucou G., 2005, « Profiler pour mieux apparier : activité des sites internet et transformation des offres d'emploi », in Marchal E. et M.-C. Bureau (éds), *Au risque de l'évaluation*, Lille, éditions du Septentrion.
- Thévenot L., 1986, « Les investissements de forme », in Thévenot (éd.), *Conventions économiques*, Paris, PUF, Centre d'études de l'emploi.

⁹ En s'appuyant récemment sur l'analyse d'offres d'emploi pour développer la strate « compétences » du Rome, Pôle emploi a fait un premier pas dans cette direction.

Les actualités du Centre d'études de l'emploi et du travail (dernières publications, colloques et séminaires) sont en ligne sur le site : www.cee-recherche.fr

Elles sont également disponibles *via* la lettre électronique flash.cee, ainsi que sur le compte Twitter  [@CeeEtudesEmploi](https://twitter.com/CeeEtudesEmploi).

Centre d'études de l'emploi et du travail

29, promenade Michel Simon - 93166 Noisy-le-Grand Cedex

Téléphone : 01 45 92 68 00 - Mèl : cee@cee-recherche.fr - site : www.cee-recherche.fr

Directrice de publication : Christine Erhel - Rédactrice en chef : Marie-Madeleine Vennat

Conception technique et visuelle : Horizon - Imprimerie : Horizon C.P.P.A.P. : 0911 B 07994 - Dépôt légal : 1705-005 - Mai 2017 - ISSN : 1767-3356