

Robots et journalistes, l'info data-driven

Description

Sur le web, les robots assurent près des deux tiers du trafic comme moteur de recherche, curateur, *spammer* et pirate. Dans la vie réelle, ils sont champion d'échecs, chauffeur, assistant chirurgien... Alors, pourquoi pas journaliste ? Outils d'une écriture « pilotée par les données » (*data-driven*), les robots arrivent dans les rédactions.

Depuis les premiers essais concluants, illustrés notamment par les performances du programme informatique Stats Monkey, inventé par l'Infolab de l'université Northwestern (Evanston et Chicago, Illinois) et capable de générer en deux secondes le compte rendu d'un match de base-ball ([voir REM n°17, p.54](#)), les robots sont sortis des laboratoires pour devenir suppléants non seulement des communicants mais aussi des journalistes.

[pullquote]**Le 17 mars 2014, c'est un robot qui, le premier, annonça qu'un séisme venait de toucher Westwood**[/pullquote]

Le 17 mars 2014, c'est un robot qui, le premier, annonça qu'un séisme venait de toucher Westwood, un quartier de Los Angeles. Présentée comme un « post *rédigé par un algorithme créé par l'auteur* », l'information fut publiée immédiatement sur le site du *Los Angeles Times* par le journaliste et concepteur de l'algorithme Quakebot, Ken Schwencke. A partir des données fournies par le Centre fédéral d'information sur l'activité sismique (US Geological Survey), le programme informatique de Ken Schwencke a fabriqué automatiquement, en trois minutes, un compte rendu mis en forme avec titre et carte à l'appui, relatant les circonstances du tremblement de terre. L'algorithme Quakebot peut de la même façon servir au traitement des faits divers, une fois programmé pour trier les données recensées dans un fichier des arrestations fourni par la police de Los Angeles.

Depuis le 21 juillet 2014, l'agence de presse américaine Associated Press (AP) publie des brèves sur les résultats financiers des entreprises, rédigées par des robots allant piocher les informations contenues dans la banque de données Zacks Investment Research. AP utilise la technologie développée par Automated Insights, entreprise américaine dont elle est actionnaire, spécialisée dans la production automatisée de contenus à partir de données brutes, principalement axée sur le sport et la finance. Mentionné comme étant « *automatiquement généré par Automated Insights utilisant les données de Zacks Investment Research* », chaque article est vérifié par un journaliste, mais cette validation devrait être, à terme, abandonnée pour la publication des informations financières de la plupart des sociétés, à l'exception des plus grandes. Si la rédaction d'un article court de 150 à 300 mots demande à peu près le même temps à un robot qu'à un journaliste, AP ambitionne d'augmenter fortement sa productivité en passant de 300 à 4 400 articles par

mois grâce aux machines. L'objectif ne serait pas pour AP de diminuer les effectifs, mais de permettre aux journalistes de se consacrer davantage à leur travail d'analyse et d'enquête. D'autres médias, tels le quotidien *USA Today* et Yahoo! News, expérimentent la génération automatique de textes. En 2013, l'algorithme Wordsmith (« manieur de mots ») d'Automated Insights a généré 300 millions d'articles ; il atteindra le milliard fin 2014, selon son fondateur et PDG Robbie Allen.

Le site web du magazine américain *Forbes* fait déjà appel, pour la publication d'informations financières, au service d'une entreprise concurrente baptisée Narrative Science. Créée en 2010, cette entreprise de Chicago, qui a déjà collecté 20 millions de dollars, commercialise le système Quill (« plume »), issu du Laboratoire d'information intelligente de l'université Northwestern. Elle compte aujourd'hui une vingtaine de clients parmi lesquels des sites d'information spécialisés dans les événements sportifs locaux ou destinés à la jeunesse, mais aussi des grandes entreprises, des banques, des courtiers et des agences de notation. Là encore, les rapports rédigés par Quill sont relus avant d'être envoyés mais, passée la phase expérimentale, cette vérification ne devrait bientôt plus être nécessaire. Fondateur de Narrative Science, le professeur Kris Hammond prédit une augmentation sans précédent du volume de textes publiés à l'avenir, précisant que « 90 % des informations lues par le grand public seront générées par des robots d'ici à 2025 », néanmoins à partir de bases de données constituées par les hommes, conviendrait-il d'ajouter.

Codirecteur du Laboratoire d'information intelligente de l'université Northwestern et l'un des inventeurs du programme Quill, Larry Birnbaum enseigne à la Medill School, école de journalisme fondée en 1921 par les héritiers de Joseph Medill (1823-1899), propriétaire du *Chicago Tribune*. Journalistes, linguistes et ingénieurs en informatique sont en train de réinventer le métier avec des machines qui ne se contentent plus de collecter et d'ordonner des masses d'informations, mais qui surtout les exploitent, afin de leur donner du sens, avant d'en composer un compte rendu en langage humain. « *Could robots be the journalists of the future* », titrait le *Guardian* en mars 2014 pour présenter Guardbot, un algorithme développé en interne afin de produire des articles présentant un point de vue. « *Le résultat ne privera certainement personne de travail* » ironise Will Franklin, inventeur de Guardbot, tout en expliquant que le robot ne parvient pas à lier les différents éléments de son récit, grammaticalement correct mais peu convaincant à la lecture.

[pullquote]« *Si le médiocre médiatique est à la portée des robots, le journalisme leur échappe.* »
Jean-Claude Guillebaud, journaliste et essayiste.[/pullquote]

Depuis septembre 2014, l'Agence France-Presse expérimente, elle aussi, la rédaction automatisée, à travers sa filiale allemande Sport Informations Dienst (SID), spécialisée dans l'information sportive. Utilisant une technologie développée par la société allemande Aexea, cette expérimentation prévue pour un an vise dans un premier temps à produire automatiquement des annonces d'événements sportifs, à partir d'un calendrier préétabli et d'une base de données enrichie par les journalistes eux-mêmes. A terme, la rédaction des articles sera adaptée à la demande de chacun des clients de l'agence.

Outre les entreprises de presse, la robotisation de l'écriture intéresse également tous les secteurs dont l'activité s'appuie notamment sur la production de nombreuses documentations : des sites web de commerce

électronique qui éditent des catalogues, des fiches de produit, des guides d'utilisation ou encore des sites web d'établissements hôteliers qui, soucieux d'être bien référencés par les moteurs de recherche grâce au descriptif *ad hoc* de leurs prestations, sous-traitent la fabrication de ces documents commerciaux aux nombreuses start-up sur ce nouveau marché. Avec un coût de production inférieur de 30 % et en deux ou trois jours seulement, la parisienne Syllabs a la capacité de fournir 150 000 notules, traduites en trois langues, destinées à la communication en ligne hôtelière ; tandis qu'Edouard de Ménibus, cofondateur de LabSense elle aussi installée dans la capitale, pense déjà au lancement d'un service en ligne offrant à chacun la possibilité d'exploiter un robot rédacteur. Implantée au Texas, la société française Yseop, quant à elle, propose à ses clients – opérateurs de télécommunications, banques et autres acteurs de la finance – d'installer chez eux un système informatique produisant automatiquement des rapports d'activité en tout genre. Selon son directeur, Jean Rauscher, l'algorithme, capable de signaler une donnée manquante, remplit une fonction d'assistant : il élabore un compte rendu servant de point de départ au travail d'analyse et de commentaire apporté ensuite par un humain.

Au sein de l'Ecole des hautes études en sciences sociales (EHESS), le moteur d'intelligence artificielle Marlowe a publié, grâce à un vocabulaire perpétuellement enrichi, plus de 900 articles sur des sujets faisant l'objet de controverses au sein de la communauté scientifique tels que les OGM, le nucléaire, le climat, le gaz de schiste ou les nanotechnologies. Créé en 2012 par le sociologue Francis Chateauraynaud et l'informaticien Jean-Pierre Charriau, le robot Marlowe utilise les informations collectées dans un premier temps par un autre moteur de recherche, « mineur de données », baptisé Tiresias, qui fouille dans une sélection de sites et de bases, données mises en forme dans un second temps par Prospero, un autre moteur de recherche « analyste », qui détermine les sujets importants en repérant notamment les cooccurrences. La principale caractéristique de Marlowe réside dans sa capacité d'analyse, c'est-à-dire l'identification de ses choix, même si ce sont les chercheurs eux-mêmes qui font ensuite le tri parmi les pistes de recherche proposées par la machine. Marlowe n'a pas vocation à être commercialisé. Mais son concepteur Francis Chateauraynaud est persuadé que l'usage des robots rédacteurs va se généraliser : « *Pour certains sites et journaux gratuits, les journalistes se contentent de réécrire des dépêches d'agence et des communiqués de presse. Quand leurs patrons découvriront que ces tâches peuvent être accomplies par une machine fiable, infatigable et bon marché, ils n'hésiteront pas une seconde* ».

[pullquote]« **Chaque article aura un seul lecteur.** »
Kris Hammond de Narrative Science.[/pullquote]

Le journaliste et essayiste Jean-Claude Guillebaud en tire une perspective plus optimiste et néanmoins sarcastique : « *Si le médiocre médiatique est à la portée des robots...* », c'est-à-dire la répétition, la routine, l'obsession du quantitatif et du raisonnement par sondage, le politiquement correct ; «... *le journalisme leur échappe* », c'est-à-dire le courage, la sensibilité, l'intuition, la créativité d'un être humain. Et d'ajouter : « *Algorithmes et logiciels peuvent bien se liquer, s'associer et se perfectionner, ils ne pourront jamais rendre compte de cette palpitation étrange – et magnifique – qu'on appelle la vie* ».

Le chercheur Kris Hammond de Narrative Science prédit, quant à lui, un bel avenir à Quill lorsque cet

algorithmes sera associé aux techniques de traçage des données d'achat, de communication, de navigation sur le web, émises par chacun des milliards d'individus.

A ce moment-là, « *chaque article aura un seul lecteur* » explique-t-il en donnant l'exemple suivant : « *Dans l'avenir, votre site d'information favori [...] saura qui vous êtes, il connaîtra la marque de votre voiture et sa consommation, la distance que vous parcourez chaque semaine, le type de carburant que vous achetez, etc. Il rédigera à votre intention un article personnalisé qui indiquera exactement combien de dollars vous économiseriez si vos pneus étaient mieux gonflés* ».

[pullquote]**Un algorithme, quoiqu'on en dise, n'est jamais neutre.**[/pullquote]

En septembre 2014, IBM a annoncé le lancement d'un service en ligne *freemium* de traitement de données, Watson Analytics, du nom du programme informatique d'intelligence artificielle célèbre pour avoir gagné, contre des humains, une finale du jeu Jeopardy en 2011, grâce à sa parfaite compréhension du langage naturel. Accessible sur tous les terminaux, y compris les smartphones et les tablettes, Watson Analytics apporte une réponse sous forme de graphique à une question posée en langage humain, à partir des données compilées dans une base de données préalablement téléchargée. Souhaitant mettre le *big data* à la portée de tous, IBM estime que le traitement et l'analyse des mégadonnées (*hardware, software* et conseil) constitueront un quart de son chiffre d'affaires d'ici à 2016, soit 20 milliards de dollars par an.

Grâce au *big data* et à l'intelligence artificielle, le « robot » et le journaliste se retrouvent à la croisée des chemins tracés par le numérique. A l'avenir, l'information sera le fruit de leur collaboration, pacifique ou non, puisqu'un algorithme, quoiqu'on en dise, n'est jamais neutre.

Sources :

- « *Could robots be the journalists of the future* », Aisha Gani and Leila Haddou, generation Y takeover, theguardian.com, 16 March 2014.
- « *Articles don't write themselves... or do they ?* », Will Franklin, developer blog, theguardian.com, 17 March 2014.
- « *The First News Report on the L.A. Earthquake Was Written by a Robot* », Will Oremus, Slate.com, 17 March 2014.
- « *Robot-journalisme : des algorithmes travaillent pour le «LA Times»* », bigbrowser.blog.lemonde.fr, 18 mars 2014.
- « *Le boom des robots de l'information* », Alice Antheaume, W.I.P., slate.fr, 7 mai 2014.
- « *Robots. Les algorithmes prennent la plume* », Yves Eudes, Science & Médecine, *Le Monde*, 2 juillet 2014.
- « *Les robots d'Associated Press ont écrit leurs premiers articles* », Thomas Oliveau, Le Figaro.fr, 21 juillet 2014.
- « *Alerte aux robots-journalistes !* », Jean-Claude Guillebaud, *CinéTéléObs*, n° 2595, 2 août 2014.
- « *Et maintenant, des robots journalistes !* », Natacha Tatu, tempsreel.nouvelobs.com, 22 août 2014.
- « *L'AFP se met à l'écriture automatique, via sa filiale allemande SID* », Stratégies.fr, 4 septembre 2014.

- « IBM mise sur la mégadonnée à la demande », Julien Dupont-Calbo, *Les Echos*, 17 septembre 2014.

Categorie

1. Articles & chroniques

date créée

9 décembre 2014

Auteur

francoise