

LE «JOURNALISME ARTIFICIEL» EST EN LIGNE

LE 14 NOVEMBRE 2010 MARC MENTRÉ

Des robots qui "rédigent" les contenus d'un journal numérique, c'est une réalité permise par les datas. StatSheet est un site d'infos sportives US où des machines intelligentes compilent des données pour en faire des articles cohérents.

C'était annoncé, c'est désormais chose faite. Les premiers articles entièrement rédigés par des « robots » sont en ligne sur le site de **StatSheet**. Pour l'instant, ces « journalistes artificiels » ne couvrent que le sport, mais Robbie Allen, le fondateur de cette *startup* américaine estime que cette technologie pourrait être utilisée à l'avenir dans d'autres domaines.

Dans un **article** publié par *Le Monde* en mars 2009, Yves Eudes décrivait *Stats Monkey*, un programme d'intelligence artificielle développé par un laboratoire d'intelligence artificielle (*infolab*), rattaché à l'université de Northwestern, à Chicago:



Il [Stats Monkey] travaille automatiquement de A à Z. Il commence par télécharger les tableaux chiffrés publiés par les sites Web des ligues de base-ball, et collecte les données brutes : score minute par minute, actions individuelles, stratégies collectives, incidents... Puis il classe cette masse d'informations et reconstruit le déroulé du match en langage informatique. Ensuite, il va puiser son vocabulaire dans une base de données contenant une liste de phrases, d'expressions toutes faites, de figures de style et de mots-clés revenant fréquemment dans la presse sportive. Il va alors rédiger un article, sans fautes de grammaire ni d'orthographe. Il peut fournir plusieurs versions, rédigées dans un style plus ou moins imagé.



À l'époque, l'article avait suscité une certaine émotion. Pourtant, depuis déjà longtemps, dans l'information financière notamment, les premiers jalons de ce type de synthèse sont posés. C'est le cas à *Bloomberg*, comme l'expliquait déjà 1999 son fondateur Michael Bloomberg:



Pour certains sujets, nous ne faisons appel ni à des journalistes, ni à des rédacteurs humains. Quand nous décrivons la valeur d'un marché à un moment donné (et non les raisons qui l'ont amené à cette situation), les deux seules choses qui comptent sont la vitesse et la précision — or ce ne sont pas les qualités essentielles de la plupart des gens. (...) Nous avons donc programmé nos ordinateurs pour qu'ils « écrivent » périodiquement une série d'articles informant nos lecteurs de l'état actuel du marché. Par exemple, la machine prend le début de la phrase « L'indice industriel Dow Jones est » et y ajoute « en hausse » ou « en baisse » en fonction du résultats de calculs portant sur les mouvements de 30 valeurs entre la veille et la micro seconde où le texte s'écrit. Elle y ajoute ensuite le chiffre adéquat, par exemple de 1 point, 2 points, 3 points, etc. Puis elle imprime: « les titres les plus actifs sont » et à partir du suivi en continu du volume des échanges, elle traduit automatiquement les symboles des téléscripteurs en noms d'entreprises (Procter & Gamble, General

Electric, Walt Disney, etc.) et les ajoute à la phrase.

[in Bloomberg par Bloomberg, Village Mondial, 1999, pages 87-88]



Le bot-style

Avec **StatSheet**, en tout cas, le pas est franchi [*le site semble encore instable*]. Le « journalisme artificiel » sort du laboratoire et **les premiers articles rédigés par des algorithmes sont en ligne, depuis août.**

Schématiquement, le procédé est le suivant, explique Allan Maurer de **TechJournal South**: *StatSheet* a stocké quelque 500 millions de statistiques, 10.000 données significatives et 4.000 phrases clé. « Les articles [*de vingt types différents*] sont entièrement auto-générés, explique Robbie Allen, son fondateur. la seule implication humaine est la création de l'algorithme qui permet de générer les articles ». [*cité par TechCrunch*]

Le résultat est surprenant. Le style est rugueux, direct et les articles bourrés de chiffres et de statistiques. « 70% du contenu sportif est basé sur des statistiques, explique Allen. Notre technologie passe par les stats, permet de faire une tonne d'analyses, et de les injecter dans des articles que l'on peut publier [*Allen dit "split" - "cracher"*] rapidement ». Voici un exemple:



Michigan State basket ouvrira la saison 2010-2011 contre Eastern Michigan le 12 novembre à East Lansing. Les attentes sont élevées pour les Spartans qui ont réalisé une excellente performance la saison dernière. Ils ont conservé 72% de leurs joueurs de la saison dernière. Ils ont complété [leurs effectifs] avec 3 recrues parmi les 100 meilleures et un étudiant de première année... [lire la version originale ici]



Le site ne se réduit pas à cette seule innovation technologique. **Il offre un contenu que ne peut pas couvrir un média traditionnel**, puisqu'il s'agit en fait d'un portail qui assure la couverture de l'actualité de 345 équipes de basketball américaines, de 1ère division ou


universitaires. Chacune d'elle bénéficie aussi d'un compte Twitter, d'une page Facebook et d'une application mobile, offrant aux supporters une couverture complète [au moins en matière de statistiques] de leur équipe favorite.

Il est probable que le « journalisme artificiel », dont on voit bien avec l'exemple de *StatSheet* les limites, ne se substitue pas au « journalisme traditionnel », mais qu'il le complète, en le libérant des tâches mécaniques comme la compilation de statistiques, et l'enrichisse. Sur ce point, on ne peut que reprendre l'affirmation de Michael Bloomberg : « Dès le début, nos journalistes ont été trop précieux pour se voir confier des tâches mécaniques ».



Crédit photos cc Flickr : **Don Solo, Gastev**.

—

Article initialement publié sur **Media Trend**.

LULU
le 14 novembre 2010 - 14:17 • SIGNALER UN ABUS - PERMALINK 

*hum, spLit ce serait plutôt découper..
cracher c'est spit*

VOUS AIMEZ  0 VOUS N'AIMEZ PAS  0 LUI RÉPONDRE

MARCMENTRE
le 14 novembre 2010 - 16:20 • SIGNALER UN ABUS - PERMALINK 

@lulu erreur (humaine) de ma part -:

VOUS AIMEZ  0 VOUS N'AIMEZ PAS  0 LUI RÉPONDRE

2 pings

3 formes de créativité machine (l'abime sans conscience) | singularite le 3 août 2011 - 20:59

[...] Le «journalisme artificiel» est en ligne [...]

Remplaçons les commerciaux | singularite le 9 février 2012 - 9:44

[...] Le «journalisme artificiel» est en ligne Détournement , et mémétique : [...]