Manière de voir
80 BOULEVARD AUGUSTE-BLANQUI
75707 PARIS CEDEX 13 - 01 57 28 20 00

FEV/ MARS 14
Bimestriel
OJD : 23612

Surface approx. (cm²) : 3105
N° de page : 62-67

des corrélations entre des dérèglements mineurs et des maux plus sérieux. Cette technique devrait permettre aux médecins d'intervenir en amont pour sauver des vies. Au fil du temps, l'enregistrement de ces observations pourrait également les aider à mieux comprendre ce qui provoque de telles infections. Cependant, lorsque la vie d'un nourrisson est en jeu, il est plus utile d'anticiper ce qui pourrait se produire que de savoir pourquoi.

L'application médicale illustre bien cette possibilité d'identifier des corrélations, même lorsque les causes qui les sous-tendent demeurent obscures. En 2009, des analystes de Google ont publié dans la revue *Nature* un article qui a fait sensation dans les milieux médicaux (1). Ses auteurs affirmaient qu'il était possible de repérer les foyers de grippe saisonnière à partir des archives du géant de l'Internet. Celui-ci gère pas moins d'un milliard de requêtes par jour sur le seul territoire américain, et conserve scrupuleusement trace de chacune de ces opérations. Il a sélectionné les cinquante millions de termes les plus fréquemment saisis sur son moteur de recherche entre 2003 et 2008, puis les a croisés avec le fichier de la grippe des centres pour le contrôle et la prévention des maladies (Centers for Disease Control and Prevention, CDC). Objectif : découvrir si la récurrence de certains mots-clés coïncidait avec les apparitions du virus ; en d'autres termes, évaluer la possible corrélation entre la fréquence de certaines recherches sur Google et les pics statistiques enregistrés par les CDC sur une même zone géographique. Ceux-ci recensent notamment les consultations hospitalières des malades de la grippe à travers tout le pays, mais ces chiffres brossent un tableau souvent en décalage d'une semaine ou deux : une éternité dans le contexte d'une pandémie. Google, lui, peut fournir des statistiques en temps réel.

La société ne disposait d'aucun élément pour deviner quels mots-clés pouvaient fournir une indication probante. Elle s'est contentée de soumettre tous ses échantillons à un algorithme conçu pour calculer leur corrélation avec les attaques du virus. Son système a ensuite combiné les termes retenus pour tenter d'obtenir le modèle le plus fiable. Après cinq cents millions d'opérations de calcul, Google est parvenu à identifier quarante-cinq mots-clés – comme « mal de tête » ou « nez qui coule » – dont la réitération recoupait les statistiques des CDC. Plus leur fréquence était grande sur une zone donnée, plus le virus faisait de ravages sur ce même périmètre. La conclusion peut paraître évidente mais, à raison d'un milliard de recherches par jour, il aurait été impossible de l'établir par d'autres moyens.

Les informations traitées par Google étaient pourtant imparfaites. Dans la mesure où elles avaient été saisies et stockées à bien d'autres fins que l'altruisme sanitaire, fautes de frappe et phrases incomplètes pullulaient. Mais la taille colossale de la banque de données a largement compensé sa nature brouillonne. Ce qui en ressort n'est qu'une simple corrélation. Elle ne livre aucun indice sur les raisons qui ont poussé l'internaute à effectuer sa recherche. Était-ce parce qu'il avait la fièvre lui-même, parce qu'on lui avait éternué au visage dans le métro, ou encore parce que le journal télévisé l'avait rendu anxieux ? Google n'en sait rien, et peu lui chaut. Il semble d'ailleurs qu'en décembre dernier son système ait surestimé le nombre de cas de grippe aux États-Unis. Les prévisions ne sont que des probabilités, jamais des certitudes, surtout lorsque la matière qui les alimente – des recherches sur Internet – est de nature aussi mouvante et vulnérable aux influences, en particulier médiatiques. Reste que les données de masse peuvent identifier des phénomènes en cours.

La transformation d'une paire de fesses en un bouquet de données numériques représente un service appréciable.

Nombre de spécialistes assurent que leur utilisation remonte à la révolution numérique des années 1980, lorsque la montée en puissance des microprocesseurs et de la mémoire informatique a rendu possibles le stockage et l'analyse de données toujours plus pléthoriques. Ce n'est vrai qu'en partie. Les progrès technologiques et l'irruption d'Internet ont certes contribué à réduire les coûts de la collecte, du stockage, du traitement et du partage des informations. Mais les données de masse constituent surtout la dernière manifestation en date de l'irrépressible désir humain de comprendre et de quantifier le monde. Pour sonder la signification de cette étape nouvelle, il faut jeter un regard de côté – ou plutôt, vers le bas.

(1) Jeremy Ginsberg, Matthew H. Mohebbi, Rajan S. Patel, Lynnette Brammer, Mark S. Smolinski et Larry Brilliant, « Detecting influenza epidemics using search engine query data », *Nature*, n° 457, Londres, 19 février 2009.

Bibliographie

Eric Heilmann, Philippe Melchior, Anne-Cécile Douillet et Séverine Germain, *Vidéo-surveillance ou vidéo-protection ?*, Le Muscadier, Paris, 2012. Confrontant les points de vue du chercheur Eric Heilmann, spécialiste de la question, et de Philippe Melchior, chargé du plan de développement de la vidéosurveillance sous la présidence Sarkozy, l'ouvrage fait le tour des enjeux soulevés par cet outil et en montre les limites. Il conclut que *« même les études les plus rigoureuses ne permettent pas toujours de tirer des conclusions définitives sur l'opportunité de la vidéosurveillance »*.

Stephen Graham, *Villes sous contrôle. La militarisation de l'espace urbain*, La Découverte, Paris, 2012. Selon les spécialistes chargés d'élaborer de nouvelles stratégies de maintien de l'ordre dans les métropoles, celles-ci formeront les principaux terrains d'engagement dans les guerres futures, dites *« de quatrième génération »*. GPS, passeports biométriques, drones, puces RFID, barrières en Jersey (en béton), etc. : les dispositifs high-tech de surveillance, d'identification et de neutralisation envahissent le quotidien des citadins, transformés de facto en ennemis potentiels.

David Forest, *Abécédaire de la société de surveillance*, Syllepse, Paris, 2009. Avocat spécialisé dans les technologies de l'information, l'auteur analyse dans le détail les mécanismes de surveillance (tests ADN, biométrie, caméras, cyberflicage, fichiers de police, géolocalisation, etc.) utilisés par les institutions et les entreprises. Dressant l'image d'une société contemporaine qui fait du contrôle systématique des individus une de ses armes majeures, au détriment des libertés individuelles, il en appelle à un sursaut citoyen.

Lawrence Lessig, *Code, and Other Laws of Cyberspace*, Basic Books, New York, 1999. Quelles sont les forces qui régissent Internet ? Il y a quinze ans, le célèbre juriste américain, spécialiste de la propriété intellectuelle et fondateur du Center for Internet and Society, prédisait qu'une coalition regroupant États et entreprises serait à même de policer le réseau en transformant son architecture de manière à en identifier, de bout en bout, les utilisateurs…

« Sociologie des bases de données », *Réseaux*, vol 31, n° 178-179, La Découverte, Paris, mai-juin 2013. Déluge numérique, vie privée, réseaux sociaux, surveillance : omniprésentes, les bases informatiques sont au cœur de ce numéro de la revue de l'université Paris-Est Marne-la-Vallée. Avec, en exemples, l'Amérique des années 1960, la santé, la cartographie, l'entreprise, etc.