

## 5. Le service reCAPTCHA

Pour certaines opérations, on vous demande de retranscrire quelques symboles mal écrits formant un mot. Le but est d'empêcher un programme de se substituer à vous et, par exemple, de créer massivement des comptes de courrier électronique. On suppose que seul un être humain peut traduire le dessin présenté en symboles. Comme ce sont des tests pour distinguer homme

et machine, tel celui qu'Alan Turing proposa en 1950, on les nomme CAPTCHA, de l'anglais *Completely Automated Public Turing Test to Tell Computers and Humans Apart*.

Une utilisation des CAPTCHA pour mettre en place un « crowd-computing » massif est pratiquée depuis plusieurs années. Le principe, nommé reCAPTCHA, consiste à demander à la personne qui s'au-

thentifie non pas une seule lecture de lettres brouillées, mais deux. Le premier mot sert à tester si l'on a bien affaire à un humain ; la seconde est une image tirée d'un document que les gestionnaires du reCAPTCHA veulent numériser et que les algorithmes de reconnaissance automatique de caractères ont repéré comme délicat ou totalement illisible. La seconde réponse de l'utilisateur permet de numériser le morceau incompréhensible, travail contrôlable par un autre utilisateur pour que la sûreté soit accrue. Ainsi sont traités des documents trop difficiles pour les numérisations automatiques. Le système est utilisé pour les archives du *New York Times*.



reCAPTCHA

nouvelle. En 2006, Jeff Howe, rédacteur au magazine *Wired*, l'a nommée *crowdsourcing* [« externalisation vers la foule » ou « distribution du calcul à grande échelle »]. Des exemples en dehors des sciences ont été mis en œuvre, dont un programme particulièrement étonnant, nommé reCAPTCHA, où les internautes participent à la validation des numérisations d'ouvrages et travaillent sans même en avoir conscience.

### Quand la foule calcule

En réfléchissant à ce *crowdsourcing*, certains ont noté des utilisations anciennes du principe datant d'une époque où les ordinateurs n'existaient pas. Ainsi, le *Oxford English Dictionary* a été élaboré au milieu du XIX<sup>e</sup> siècle avec l'aide de 800 volontaires qui copiaient des passages de livres, ce qui mit à la disposition des éditeurs un grand nombre de citations précises et une information fiable sur l'usage des mots et leur sens en langue anglaise. *Wikipedia* est bien sûr une forme de *crowdsourcing* et, bien qu'écrite par ceux qui le souhaitent, sa fiabilité n'est pas loin d'égaler celle des autres encyclopédies, tout en proposant une quantité d'articles bien supérieure à

la plus grosse des encyclopédies fondées sur le travail d'experts.

Notons toutefois que le fait de confier une tâche à un très grand nombre de personnes ne produit pas nécessairement des résultats meilleurs que ceux qu'on tire d'un individu isolé et compétent. L'exemple le plus spectaculaire d'un échec du *crowdsourcing* est donné par le match organisé en 1999 entre G. Kasparov et le reste du monde, constitué de 50 000 personnes provenant de 75 pays différents. Durant toute une partie qui s'étendit sur 62 coups, à raison d'un coup tous les deux jours, G. Kasparov affronta la foule des joueurs d'échecs du monde entier qui tentait de le battre. Malgré une organisation permettant aux bons joueurs de peser suffisamment dans les décisions collectives des coups à jouer, G. Kasparov gagna. La partie a été jugée l'une des plus importantes et des plus riches de l'histoire du jeu d'échecs, et un commentaire de 200 pages a été publié pour en analyser les nombreuses nouveautés et subtilités. Reste que la foule a perdu ! N'oublions pas non plus que la « sagesse des masses » a conduit aux plus grandes bulles spéculatives, à de multiples catastrophes économiques, et à l'avènement des pires dictateurs. ■

## L'AUTEUR



Jean-Paul DELAHAYE est professeur à l'Université de Lille et chercheur au Laboratoire d'informatique fondamentale de Lille (LIFL).

## BIBLIOGRAPHIE

A. Kawrykow et al., *Phylo : A citizen science approach for improving multiple sequence alignment*, *PLoS ONE*, vol. 7(3), e31362, 2012.

Kasparov versus the world, *Wikipedia*, 2012 : [http://en.wikipedia.org/wiki/Kasparov\\_versus\\_the\\_World](http://en.wikipedia.org/wiki/Kasparov_versus_the_World)

S. Thomas, *9 examples of crowdsourcing, before crowdsourcing existed*, 2012 : <http://memeburn.com/2011/09/9-examples-of-crowdsourcing-before-crowdsourcing-existed/print/>

F. Khatib et al., *Algorithm discovery by protein folding game players*, *PNAS*, vol. 108(47), pp. 18949-18953, 2011.

F. Khatib et al., *Crystal structure of monomeric retroviral protease solved by protein folding game players*, *Nat. Struct. Mol. Biol.*, vol. 18, pp. 1175-1177, 2011.

C. Lintott et al., *Galaxy Zoo : Morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey*, *Monthly Notices of the Royal Astronomical Society*, vol. 389(3), pp. 1179-1189, 2008.

J. Howe, *The rise of crowdsourcing*, *Wired Magazine*, juin 2006.

J. Surowiecki, *The Wisdom of Crowds*, Anchor Books, 2004.