

ENTRETIEN

« La connaissance n'est pas réductible à l'information »

Cette année 2016 marque le centenaire de la naissance de l'Américain Claude Shannon, l'un des pères de la théorie de l'information : l'occasion de revenir sur ce terme abondamment employé... et sur ses limites.



Jean-Louis ERMINE est professeur émérite à l'institut Mines-Télécom et expert en gestion des connaissances.

POUR LA SCIENCE

Pouvez-vous rappeler comment la théorie de Shannon définit l'information ?

JEAN-LOUIS ERMINE : La notion forgée par Shannon dans le cadre de sa théorie des télécommunications est assez éloignée de la signification courante du mot information. On considère une source qui émet un message, codé à l'aide de signaux élémentaires, à travers un canal de communication, et un récepteur qui doit décoder ce message. Supposons que le récepteur ait N décodages possibles du message, chacun ayant une certaine probabilité. La quantité d'information ou entropie de Shannon du message est alors, par définition, la quantité $-(p_1 \log p_1 + \dots + p_N \log p_N)$, où p_i est la probabilité associée à la i -ème possibilité et où le logarithme est, par convention, celui de base 2.

PLS

Quel est le rapport avec les bits d'information ?

J.-L. E. : Le bit est l'unité de mesure de l'information au sens ci-dessus. Or les techniques informatiques utilisent des signaux binaires représentés par 0 ou 1 et qu'on appelle aussi bits. Si les signaux 0 et 1 sont de même probabilité (égale à $1/2$), alors l'information correspondant à un bit informatique

est égale, selon la définition de Shannon, à $-(1/2) \log(1/2) - (1/2) \log(1/2)$, ce qui fait 1 bit d'information (car $\log(1/2) = -1$).

PLS

L'information à la Shannon est-elle un concept satisfaisant ?

J.-L. E. : En informatique et dans les technologies de la communication, oui. Mais l'information à la Shannon, ce n'est pas l'information du 20 heures à la télévision ! Elle ne mesure pas du tout le contenu sémantique des messages, c'est-à-dire leur signification. Elle ne rend donc pas compte des nombreuses acceptions qu'a, dans le langage courant, le mot information. Beaucoup de chercheurs ont essayé d'élaborer des définitions de l'information au sens sémantique qui soient précises et si possible quantitatives. Après des décennies de recherches et de polémiques, on arrive à peu près à un consensus appelé GDI (*General Definition of Information*), que l'on peut résumer en disant qu'une information est un ensemble structuré de données dotées d'un sens pour les acteurs du domaine concerné.

PLS

Pouvez-vous préciser la différence entre données et informations ?

J.-L. E. : Une donnée désigne un fait brut, une observation, une perception, un signal de

capteur... Par exemple, on peut avoir pour donnée le nombre 454 ; ce nombre devient une information lorsqu'on dit qu'il est la valeur de l'indice boursier CAC 40 à telle date. En d'autres termes, pour que des données deviennent des informations, il faut les traiter pour leur donner du sens. Les bases de données sont ainsi plutôt, à proprement parler, des bases d'informations : les données ne sont pas seulement rassemblées, elles sont aussi classées d'une certaine façon, complétées par des précisions, ce qui fait d'elles des informations, élémentaires certes, mais des informations quand même.

PLS

Vous êtes un spécialiste de gestion des connaissances. Comment s'articulent les notions d'information et de connaissance ?

J.-L. E. : Une connaissance est une information qui prend du sens dans un contexte donné. Elle n'est pas réductible à l'information. Voici deux exemples simples pour mieux comprendre. Les bases de données permettent parfois de conférer du sens à leurs contenus, mais elles ne fournissent pratiquement jamais de contexte d'interprétation ; elles produisent des informations (rudimentaires), mais pas de connaissances. C'est un peu différent avec le moteur de recherche de Google. Lorsqu'on lui soumet