L'information grise

[analyse] Appréhender la notion d'information grise conduit à aborder la question éthique. Celle-ci monte en puissance avec l'internet et le rôle joué par les réseaux sociaux où l'humain garde toutefois une place essentielle. Cet article clarifie cet aspect incontournable de la veille.

Frédéric MARTINET exerce depuis 15 ans dans des activités liées à la veille et à l'intelligence économique. En 2009, il crée le cabinet de conseil indépendant Actulligence Consulting qui accompagne les entreprises dans la mise en place ou les évolutions de leurs systèmes de veille. Depuis 2001, il partage son expérience de la veille sur le site actulligence.com.

frederic.martinet@actulligence.com

L'information grise est généralement définie comme une information accessible de façon légale mais « *caractérisée par des difficultés dans la connaissance de son existence ou de son accès* »[1]. Il s'agit donc d'une notion qui complète celle de littérature grise (considérée en partie comme une information grise) mais qui, ayant bénéficié des facilités apportées par Internet pour la mise à disposition d'informations non commercialisées, devient une information blanche.

Souvent associée à l'« intelligence » dans son acception anglophone de renseignement et, plus particulièrement, de renseignement terrain, l'information grise est une information complexe, polymorphe, aux canaux de diffusion hétérogènes. Elle peut être acquise de façon licite, mais aussi quelquefois éthiquement discutable. Elle peut être implicite, déduite, calculée, avec un indice de confiance variable en fonction des sources primaires ayant servi à sa construction.

L'éthique de l'accès à une information

L'éthique est difficile à cerner car souvent variable au fil du temps et d'un interlocuteur à un autre. Au final, elle se définit souvent par ses infractions. Dans un environnement économique auquel certains associent le terme de « guerre », on peut imaginer que l'asymétrie de position dans l'accès à l'information entre deux concurrents, dont l'un serait éthique et l'autre non, se traduise, pour le plus intègre, à courir un 100 mètres haies avec une paire de chaînes. La meilleure façon de protéger son information grise serait donc de la mettre en terrain noir.

Parmi les actions connues comme étant éthiquement discutables, on peut citer : faire les poubelles de ses concurrents ayant oublié de déchiqueter des documents, prendre sa pause cigarette au pied de l'immeuble de son concurrent aux heures d'affluence, etc. Que feriez-vous enfin si l'un de vos concurrents ouvrait son ordinateur à côté de vous, dans un avion par exemple ? Déclineriez-vous votre identité et votre fonction ?
Inattentions, erreurs, oublis sont autant de paramètres permettant d'accéder à l'information grise.

Information grise et réseaux sociaux

L'information grise se caractérise par sa difficulté d'accès ou même d'identification. En ce sens, les réseaux sociaux sont indiscutablement une source d'information grise. Facebook ou LinkedIn, par exemple, regorgent d'espaces que l'on qualifiera de « semi-privatifs »

car, à moins d'être naïf, il est illusoire de penser qu'une information numérisée, mise à la disposition d'une société privée et partagée avec ses « amis » soit privée. Qu'il s'agisse de groupes privés sur LinkedIn ou de vos statuts partagés avec vos proches, voici autant de sources d'information qui ont relégué les techniques de *social engineering* traditionnelles aux oubliettes.

Google indexe tout le Web

Non, bien sûr, et la récente disposition sur le droit à l'oubli en donne un exemple : pour nos politiciens, une information non trouvée sur Google est une information qui n'existe pas (à moins d'utiliser la version états-unienne de Google non soumise au même droit). En France, on considérera (assez justement) qu'une information non disponible sur Google est une information grise car difficilement accessible. Or, Google est très loin d'indexer tout le Web. Sans parler des espaces sécurisés, il suffit de citer :

- les fichiers robots.txt interdisant à Google d'indexer certaines pages Web ou encore les méta-balises robots *no follow/no index*,
- les fichiers de formats non ou mal indexés tels que le Flash,
- les contenus multimédias indexés uniquement *via* le contenu textuel de la page et non *via* la piste audio et encore moins vidéo (même si, avec Google Audio Indexing, Google s'était essayé au *speech to text*),
- les contenus interdits d'indexation par les autorités locales,
- les serveurs hébergeant du contenu « *dark Web* » (piratage logiciel, pornographie et trafics en tout genre),
- les contenus protégés par le Digital Millenium Copyright Act aux États-Unis.

En dehors de ces freins, un cas d'information grise est encore plus difficile à traiter. En effet, même si un document comporte le mot clé recherché, Google ne vous le renverra pas systématiquement dans les résultats, car il peut considérer que sa présence est anecdotique.

L'information déduite ou calculée

En octobre 2013, les chercheurs de Facebook ont démontré qu'ils pouvaient deviner, à partir des échanges (*likes*/commentaires) entre les membres, qu'une relation de couple était sur le point de se terminer.

La donnée n'existe pas. Elle est calculée. Nous avons tous eu l'occasion de voir des CV LinkedIn se « réactiver » et deviner que la personne mentionnée passait en « écoute active » du marché. La théorie des réseaux nous enseigne que les interactions, aussi minimes soient-elles, trouvent un ancrage, un intérêt, un but poursuivi dans la vraie vie. Pour schématiser, un simple *like* est porteur de sens. Plusieurs le sont encore plus.

L'exploitation des interactions sur les réseaux sociaux (*social network analysis*) et celle des technologies de *big data* permettent de traiter des données transactionnelles ou comportementales en ligne pour anticiper les comportements ou détecter les relations non explicites.

Reconstitution d'itinéraires individuels avec les solutions de géolocalisation telles que Foursquare ou d'objets connectés, analyse sémantique et lexicale des prises de parole publiques, requêtes sur les moteurs de recherche ou pages visitées : autant de données de grande valeur si l'on se donne la peine de mettre en place les dispositifs de valorisation et de compréhension de ces informations. Les technologies actuelles permettent de réaliser le fantasme du veilleur et de faciliter la détection et la vérification du signal faible.

Le poids du PBCK

Une stratégie de sécurisation de l'information par les entreprises se limitant à la sécurité informatique n'est plus suffisante. La protection de l'information passe par la sensibilisation des personnels à sa valeur et aux règles de prudence élémentaires. « *The problem is between chair and keyboard* » (PBCK) - en somme, le problème, c'est le facteur humain.

A contrario, pour celui en quête d'information grise, fréquenter les mêmes espaces, les mêmes festivités alcoolisées que ses concurrents est un facteur clé de succès quand il s'agit d'acquérir de l'information potentiellement intéressante. Là encore, les réseaux sociaux ont facilité l'accès à ces types d'information : savoir qui est à quelle soirée, à quoi il ressemble et les sujets qui l'intéressent. Un *social engineering* qui ne dit pas son nom.

Ne pas franchir la ligne jaune

Aujourd'hui, les entreprises souffrent d'une surabondance d'information ou plutôt d'une inertie à mettre en place des processus, des règles, des outils adaptés pour faire face à l'explosion des volumes. L'information grise est souvent une information blanche noyée dans une masse, et tous ceux qui recherchent des pépites grises devraient prendre soin d'exploiter ce qui est déjà à leur portée.

Mais un veilleur ne doit jamais franchir la ligne jaune le faisant basculer dans l'illégalité. Or, c'est loin d'être toujours le cas dans un univers numérique où la réglementation et la jurisprudence évoluent continuellement.

Amis veilleurs, soyez donc imaginatifs, inventifs, voire impitoyables. L'information est votre matière première : le cœur de votre métier est d'aller la chercher, de la façonner et de l'utiliser ! ■

www.adbs.fr/information-grise-17424.htm?RH=OUTILS_VOC