

La veille informationnelle

Enjeux et évolutions

La veille professionnelle est aujourd'hui confrontée à de nouveaux défis étroitement liés au développement des technologies de l'information. Décryptage par Frédéric Martinet, consultant indépendant en intelligence économique.

La veille, démarche qui vise à rationaliser, organiser et systématiser la surveillance de son environnement, et plus particulièrement des mouvements qui pourraient avoir un impact sur son activité, est inévitablement confrontée à la numérisation de l'information et à l'explosion de ses volumes et donc aux aspects techniques qu'il est désormais impossible d'éviter lorsqu'il s'agit d'accéder à l'information et la traiter.

LA VEILLE À L'HEURE DE LA NUMÉRISATION

La numérisation n'en est certes plus à ses débuts, mais c'est un mouvement massif qui vise désormais à supprimer l'information sur support physique ou tout du moins sur support papier. Cette évolution du support de l'information a transformé les métiers du savoir et déplacé l'intervention humaine dans la chaîne de traitement de la valeur. Là où l'humain lisait, sélectionnait, résumait – travail que nombre de documentalistes ont réalisé fort longtemps – il ne subsiste maintenant plus que des algorithmes qu'il utilise et paramètre en y injectant des mots-clés, des périodes, en peaufinant des équations booléennes et en couplant son système avec des sources d'information.

La veille requiert désormais les capacités à :

- détecter de nouvelles sources d'informations ;
- se connecter à ces sources afin d'en extraire l'information de la façon la plus efficace possible ;
- savoir utiliser des outils de traitement de ces flux d'information et maintenir leur efficacité ;
- analyser cette information ;
- animer une communauté de personnes mobilisées autour d'un processus de veille.

LES NOUVEAUX DÉFIS DES ENCLOSURES¹

Il y eut d'abord le Web 1, un idéal ouvert de documents liés les uns aux autres par une balise unique : la balise <A>. Puis vint le Web 2, un web indéniablement social, favorisant l'explosion des contenus et plus particulièrement des contenus générés par les utilisateurs (*User-generated content*), avec l'apparition d'un nouvel Eldorado : celui de la donnée des utilisateurs, de la donnée personnelle et de sa monétisation.

De ce web normé, unifié, social naquirent des entités

dont le slogan désormais lointain était « *don't be evil* » (« ne soyez pas malhonnêtes »). Les univers, ouverts dans un premier temps, se referment pour céder la place à des espaces clos, privatifs ou « communautaires » qui protègent leurs données à coup de conditions générales d'utilisation (CGU), de procès ou d'API de plus en plus restrictives, et le veilleur se trouve ainsi confronté à des murs numériques.

Les options sont alors multiples et leur impact éthique ou juridique souvent loin d'être neutre :

• pénétrer ces espaces en allant à l'encontre des CGU ;

• utiliser des APIs souvent bien restrictives et perdre une part des données ou des éléments de contextualisation pouvant ainsi amener à déformer l'information ;

• acheter la donnée lorsque cela lui est possible.

Il s'agit là du défi le plus important que doit relever le veilleur car ces espaces donnent accès à de l'information blanche, grand public, massive, mais aussi à des informations parfois grises, des indiscretions, des inattentions à valeur souvent bien plus élevée.

COMPOSER AVEC LA MASSE D'INFORMATION

À lui seul Google posséderait à ce jour plus de deux millions et demi de serveurs informatiques, faisant de lui le premier possesseur privé de serveurs en termes de nombre, de capacité de stockage et probablement en capacité de calcul.

Même si le veilleur limite son attention à une partie du web (celle relative à son activité), il n'en demeure pas moins que la surveillance de millions de pages web nécessite du matériel et de la bande passante. De surcroît, même s'il limite son périmètre d'écoute, il aura toujours besoin d'avoir un radar d'horizon plus large sur le web, afin de détecter de nouvelles sources ou tout simplement des informations qui seraient diffusées sur des sources non surveillées. La masse d'information disponible et accessible pose donc plusieurs problèmes :

• celui de l'aiguille dans la botte de foin et plus particulièrement lorsque l'on recherche un signal faible ou divergent dont les mots clés descripteurs ou le vecteur sont inconnus ;

[1] Le terme « enclosure » est utilisé pour décrire des processus visant à limiter l'accès ou la réutilisation d'une information.

- celui du bruit informationnel, car quelle que soit l'efficacité des équations de recherche ou de filtre, quelle que soit la qualité des algorithmes de détournement de l'information, de désambiguïsation ou tout simplement d'anti-spam, le veilleur se heurte bien souvent au bruit ;
- celui de la mobilisation des ressources humaines, car les effectifs de veilleurs et de documentalistes sont aujourd'hui rarement à la hausse et je reste convaincu qu'à ce jour la qualité des algorithmes et des technologies accessibles au commun des entreprises ne suffit pas à pallier les limitations imposées en termes de personnels chargés de la collecte, de la diffusion et de l'analyse de l'information

LES CONTRAINTES JURIDIQUES

La veille est soumise à de fortes contraintes réglementaires, juridiques et contractuelles. Et cela ne semble pas évoluer dans un sens qui facilite les activités de veille malgré des mouvements comme les *creative commons* et l'*open data* qui en sont encore à leurs premiers frémissements.

Bien que dans les faits, de nombreuses entreprises sous-estiment, ignorent ou tout simplement décident de ne pas respecter certaines restrictions juridiques, toute information qu'elle soit papier ou numérique (site web, blog, forum...) est potentiellement soumise au droit d'auteur à de rares exceptions près. Par ailleurs, les éditeurs de sites web ajoutent souvent des conditions générales d'utilisation qui interdisent l'extraction automatisée des informations qui y sont diffusées.

La législation est également différente en fonction du type d'information, des structures, mais aussi du pays du détenteur des droits de propriété intellectuelle, de l'hébergeur et du veilleur, complexifiant souvent à outrance les veilles dont la portée est internationale et l'appréciation du risque juridique y afférent. Enfin, la jurisprudence est assez mouvante sur ces sujets et le veilleur devra avoir à cœur de se tenir informé des derniers jugements.

DES DIFFICULTÉS SPÉCIFIQUES AU MULTIMÉDIA

Rechercher des textes à partir de mots-clés devient presque une banalité à tel point que l'on entend souvent l'hérésie « *tous veilleurs* » ou bien alors « *sur Google, on trouve tout* ». Par contre, la recherche d'images et la recherche vidéo comportent encore de nombreuses carences qui limitent leur accessibilité au veilleur. Ainsi la recherche d'images à partir de mots-clés nécessite une indexation manuelle car hormis quelques formes pouvant être reconnues automatiquement ou quelques textes ocrésés, l'indexation automatique et la catégorisation des images restent encore limitées. On voit bien apparaître des technologies de recherche d'images



par similarité à une autre image, mais celles-ci sont très gourmandes en ressources informatiques et en temps de calcul limitant souvent leur exploitation à quelques sites (Flickr en tête) ou à quelques applications (monuments, étiquettes de vin, produits et marques, véhicules et plus particulièrement véhicules militaires...).

De la même façon, pour les flux audios ou vidéos, les logiciels de reconnaissance vocale (*Speech to Text*) sont également gourmands en ressources et leur exploitation reste coûteuse, limitant leur accès aux organismes gouvernementaux, aux grands groupes ou aux prestataires de services à même de mutualiser ces technologies pour leurs clients.

En conclusion, les défis de la veille sont étroitement liés à l'évolution de l'information et ses nouveaux enjeux. Là où certaines sociétés privées cherchent à s'accaparer l'information, le veilleur devra trouver moyen d'y accéder en respectant les contraintes réglementaires. Alors que les ressources humaines dédiées à la veille s'amenuisent, ce dernier devra sélectionner les technologies adéquates et mobiliser les bonnes volontés internes, les connaissances et s'affirmer comme un animateur-coordonateur. Enfin, le veilleur devra garder son esprit ouvert sur un web en mouvement perpétuel. Il évaluera les nouvelles opportunités d'accès à l'information et devra savoir composer avec les évolutions technologiques.

FRÉDÉRIC MARTINET

Fondateur d'Actulligence Consulting
 frederic.martinet@actulligence.com
 www.actulligence.com

↖ Le veilleur aura toujours besoin d'un radar d'horizon sur le web afin de détecter, dans le flux massif des informations, de nouvelles sources utiles à son activité.