

ANALYSER L'INFORMATION POUR FIABILISER LA PRISE DE DÉCISION

Véronique MESGUICH

Consultante formatrice spécialisée en veille stratégique et management de l'information

■ Qu'elle soit effectuée par des personnes humaines ou des solutions automatisées, l'analyse occupe une place cruciale dans le processus de veille. Cet article présente différentes approches et méthodes d'analyse de l'information stratégique : analyse de la qualité et de la fiabilité de l'information, identification de signaux faibles, visualisation de données... La valeur ajoutée des professionnels de l'information dans ce processus consiste ici non seulement en la maîtrise des techniques et solutions présentées, mais aussi en une interprétation judicieuse et non biaisée des résultats de la veille.

■ Analyse blijft een cruciaal element in het proces, ongeacht of ze wordt uitgevoerd door de mens, dan wel door geautomatiseerde oplossingen. In dit artikel worden verschillende benaderingen en methoden gepresenteerd voor de analyse van strategische informatie: analyse van de kwaliteit en betrouwbaarheid van de informatie, identificatie van zwakke signalen, visualisering van gegevens ... De toegevoegde waarde van de informatieprofessionals in dit proces schuilt hier niet alleen in de beheersing van de gepresenteerde technieken en oplossingen, maar ook in een oordeelkundige en onbevooroordeelde interpretatie van de resultaten van de monitoring.

Au fil des années, les méthodes et outils de veille ont évolué conjointement avec les usages et besoins des destinataires. Les objectifs classiques de la veille stratégique (surveiller les produits nouveaux et la concurrence, détecter des menaces ou des opportunités) se sont enrichis de nouveaux besoins : repérer les nouveaux entrants dans un secteur, développer l'innovation ouverte, identifier des experts et des influenceurs, identifier et cartographier des liens cachés au sein d'un écosystème, faire émerger des modèles prédictifs...

Dans ce contexte, face à des utilisateurs toujours plus connectés et capables de trouver par eux-mêmes l'information utile (mais souvent noyés dans l'infobésité), face à des solutions de collecte toujours plus automatisées, la valeur ajoutée des professionnels de l'information réside, entre autres, dans la mise en œuvre et la validation du processus d'analyse.

L'analyse occupe une place importante dans le cycle du renseignement ; c'est en effet cette phase qui va permettre de dégager les informations utiles, à partir des résultats d'une recherche ou d'une collecte automatisée, afin de mettre en évidence des tendances ou innovations. L'analyse va non seulement faciliter la compréhension d'une situation complexe, mais aussi aider à la prise de décision, voire prédire des évolutions futures. Les fameux "signaux faibles", ces signes avant-coureurs permettant d'anticiper les mutations ou de saisir les opportunités, sont difficiles à détecter par une simple lecture et résultent souvent d'une analyse automatisée.

La complexité de l'analyse, dans le domaine de la veille stratégique, provient non seulement des volumétries d'informations (données ou documents), mais aussi de leur hétérogénéité : documents textuels classiques,

documents multimédia, données structurées ou non, produites par des humains ou des objets connectés...

Cet article présente différentes approches en analyse de l'information, ainsi que les principales méthodes et outils, en nous plaçant dans le contexte actuel des pratiques de veille, à l'heure d'une production de contenus de plus en plus en temps réel et multimédia, et du développement des pratiques collaboratives. L'analyse, qu'elle soit humaine ou automatisée, peut en effet concerner la qualité et la fiabilité de l'information, mais aussi et surtout faire émerger signaux faibles et corrélations. Nous présenterons un bref panorama des solutions disponibles, notamment dans le domaine en pleine effervescence de la datavisualisation.

Analyser la qualité et la fiabilité de l'information

À l'ère des fake news, du "deep fake" et de la post-vérité, il est crucial d'analyser la qualité, la véracité et la fiabilité de l'information trouvée sur le web et les réseaux sociaux. L'analyse peut permettre de ne pas se laisser abuser par la désinformation (information intentionnellement fautive ou tronquée, visant à nuire) ou la mésinformation (information sans intérêt, redondante ou de mauvaise qualité).

Dans ce domaine, l'expertise humaine, qui reste encore supérieure aux algorithmes, se base sur plusieurs critères classiques : autorité cognitive du ou des producteurs de l'information, date de publication ou de mise à jour, clarté, cohérence et caractère innovant du fond et de la forme, objectifs et éthiques de la publication...

Plusieurs extensions de navigateurs, basées entre autres sur les annuaires de noms de domaines,

peuvent contribuer à vérifier l'origine d'un site. En revanche, la traçabilité est plus complexe lorsqu'il s'agit de blogs ou de profils sur les réseaux sociaux. Afin de lutter contre la désinformation, plusieurs initiatives ont été mises en place notamment par des médias comme la RTBF avec le site *Faky.be* ou bien Le Monde avec le service *Decodex*. Dans le domaine de l'information scientifique, les indicateurs bibliométriques contribuent à mesurer l'importance dans la sphère académique, d'un chercheur, d'une équipe ou d'un établissement de recherche, en fonction du nombre de ses publications et du nombre de citations dont il fait l'objet. Conséquences de l'évolution du cadre traditionnel de la publication scientifique, sont apparus de nouveaux indicateurs, sous le nom d'*Almetrics*, qui mesurent l'impact sur Internet d'une publication : ces nouvelles métriques viennent compléter les mesures classiques. L'analyse bibliométrique vise également, toujours dans le cadre d'une veille technologique ou d'innovation, à identifier des sujets émergents ou des partenariats de publication.

Faire émerger des signaux faibles, des informations critiques

Face aux volumes gigantesques et à l'hétérogénéité des informations à traiter, l'analyse automatisée s'impose plutôt que les modèles classiques d'analyses humaines. En effet, ces modèles d'analyse stratégique, souvent issus de grands cabinets de conseil anglo-saxons (SWOT, PESTEL, 5 forces de Porter) apparus dans la deuxième moitié du XX^{ème} siècle, présentent aujourd'hui parfois des limites face à la complexité des enjeux et des données.

Le choix entre les nombreuses méthodes, techniques et solutions d'analyse automatisée, dépendent bien entendu des besoins et des types de données à traiter : pages web, textes non structurés, contenus multimédias, données statistiques, publications issues des réseaux sociaux... Il existe de nombreuses techniques et approches, allant du clustering (catégorisation automatique) à la fouille de textes ou de données, en passant par la datavisualisation et l'analyse du sentiment et des avis exprimés sur les réseaux sociaux. Le clustering permet de classer automatiquement des objets (pages web, images, données...), en utilisant des méthodes statistiques. Les objets sont ainsi regroupés selon plusieurs catégories, définies ou non a priori. Certaines solutions de veille mettent en œuvre des technologies de clustering afin de faire émerger des tendances, ou de faire apparaître des corrélations.

La fouille de textes ou de données (text ou data mining) consiste à extraire des connaissances depuis un corpus important de documents textuels ou d'extraction de grandes bases de données. Là

encore, l'objectif est de dégager des tendances ou des corrélations entre éléments, en mettant en œuvre des traitements statistiques ou sémantiques.

La veille e-reputation utilise plutôt des algorithmes d'analyse des sentiments exprimés sur les réseaux sociaux. Il s'agit d'évaluer la tonalité d'une publication sur les médias sociaux, concernant une personne, ou une marque, en fonction d'une analyse sémantique des termes ainsi que des émoticônes positifs ou négatifs utilisés dans la publication. Il convient en revanche de prendre en compte, le cas échéant, l'ironie, la litote ou le second degré souvent présents sur les réseaux sociaux !

La datavisualisation nouvelle génération

De nombreuses solutions de veille intègrent désormais des fonctions de datavisualisation. Ce terme à géométrie variable s'applique à plusieurs cas. L'un des objectifs de la datavisualisation est de faciliter la compréhension d'un phénomène ou d'une situation, grâce à des graphiques ou des infographies, dans une logique plutôt conceptuelle. Face à des volumétries de plus en plus importantes de données hétérogènes, sont apparues des approches davantage basées sur la découverte et l'exploration : on parle de "data driven analysis", qui permet d'identifier de façon prédictive des tendances ou des influenceurs, de mettre en évidence des liens cachés ou peu apparents, de tester des hypothèses, ou d'anticiper des crises. La mise en ligne d'innombrables jeux de données ouverts (open data) et l'apparition de nouvelles solutions de "data storytelling" ont contribué à populariser et à démocratiser la datavisualisation, notamment via les pratiques de datajournalisme.

La datavisualisation met en œuvre des modèles de représentation très divers, et il convient de les choisir en fonction des types de données à analyser, et de l'objectif général. Les modèles classiques (histogramme, graphiques en courbes, nuages de points, camembert...) servent à comparer des valeurs, afficher des interactions entre valeurs, ou représenter des pourcentages. La plupart des solutions de datavisualisation offrent des modèles plus complexes, comme le diagramme "rayon de soleil", qui sert à visualiser plusieurs niveaux hiérarchiques et leurs proportions, le diagramme de Sankey destiné à la représentation de données de flux, ou encore le "chord diagram" qui représente les relations entre les données d'un tableau croisé. Plusieurs solutions de datavisualisation mettent en œuvre des graphes, c'est-à-dire une représentation de données et de leurs relations. Dans un graphe, les données sont représentées par des points (appelés sommets, ou nœuds) et les relations par des traits présentant ou non une direction. Les graphes sont

fréquemment utilisés pour la visualisation de liens entre plusieurs éléments, comme par exemple des relations entre parties prenantes, entre des termes cooccurrents au sein d'un texte, ou encore pour isoler des communautés dans les réseaux sociaux à partir de l'analyse des échanges. Une communauté sera ainsi définie en fonction de la détection d'un groupe de personnes qui échangent entre eux de façon importante. À partir d'une extraction de contenus d'un réseau social sous forme de tableau de données, un logiciel de visualisation de graphes comme *Gephi* est capable de représenter les liens existants entre différentes données. L'analyse du graphe via des algorithmes de détection facilite ainsi la compréhension, et donne du sens à des volumes de données importants.

Un large éventail de solutions

Les solutions d'analyse automatisée sont très nombreuses et diverses, à l'instar des outils de veille. Il n'existe pas de solutions miracles, et les coûts de licence peuvent parfois être onéreux. Notons également que certaines solutions, gratuites ou non, requièrent un niveau d'expertise en développement.

Les solutions de veille professionnelles, qui couvrent toutes la chaîne du processus (*Digimind, Bertin IT, Geotrend, KBCrawl...*) intègrent des outils d'analyse correspondant à plusieurs types de besoins : tableaux de bord, extraction de concepts et de nuages de mots, représentations de liens entre entités... Les plateformes d'écoute des médias sociaux (social media listening) comme *Visibrain* ou *Alerti* mettent en œuvre des techniques automatisées permettant

de détecter des influenceurs ou des communautés, et également d'analyser le sentiment exprimé dans des publications sur les réseaux sociaux.

On peut opter également pour des outils dédiés à l'analyse automatisée. Parmi les plus utilisés, citons de nombreux générateurs de tableaux de bord (*Tableau Software, Slemma, Infogr.am*), sans oublier l'outil gratuit *Google Data Studio*. Toutes ces solutions, dont certaines sont accessibles en version publique et gratuite, créent à partir des données de l'utilisateur des datavisualisations interactives et personnalisables, basées sur différents types de représentations.

Les bibliothèques javascript comme *D3js* ou *Chartjs* sont open source, mais nécessitent la maîtrise de codes informatiques pour transformer des données numériques en formes graphiques. Également open source, des logiciels d'analyse de réseaux comme *Gephi* ou *NodeXL* offrent la visualisation de grands graphes, basés sur des données créées par l'utilisateur ou issues des réseaux sociaux.

Plus accessibles aux néophytes, disponibles en version gratuite ou premium, les services en ligne de création d'infographies (*Piktochart, Canva*) ou de frises chronologiques (*Genially, Timeline Storyteller*) offrent des modèles à compléter et à enrichir. Pour des analyses de données géolocalisées, on utilisera des systèmes d'information géographiques comme *Arcgis* ou plus simplement des systèmes permettant de géolocaliser des points sur une carte géographique (*Umap/Openstreet map, Google My Maps*).

Corpus à analyser	Objectifs stratégiques	Méthodes	Livrables
Ensemble réduit d'articles et pages web	Aider à la compréhension	Analyse humaine basée sur des grilles préétablies	Étude, état de l'art, infographie, cartographie de controverses
Corpus important de documents numérisés (ex : articles scientifiques)	Faire émerger des corrélations, des signaux faibles Classer automatiquement	Fouille de texte : analyse statistique et/ou sémantique Clustering	Graphes, nuage de mots, extraction de concepts
Publications sur les réseaux sociaux	Rendre visibles des liens entre les personnes, identifier des communautés, des tendances	Analyse statistique Datavisualisation Analyse du sentiment	Tableaux de bord, graphes
Brevets	Rendre visibles des relations entre déposants, des secteurs émergents	Analyse statistique Datavisualisation	Tableaux de bord, graphes
Données ouvertes (open data)	Expliquer des phénomènes (ex : data journalisme)	Analyse statistique Cartographie de données Datavisualisation	Graphiques, représentations cartographiques
Données financières	Diagnostiquer les comptes d'une entreprise, comparer des entreprises...	Analyse financière	Graphiques

Tab. 1 : méthodes à adopter en fonction des corpus à analyser et des objectifs stratégiques

Concernant l'analyse de données textuelles, citons entre autres l'application open source *Voyant Tools*. À partir d'un texte ou d'une URL, l'application génère des nuages de tags, calcule les occurrences et cooccurrences de termes, et offre de nombreuses possibilités d'analyse automatisée de textes.

Les solutions apparues récemment sur le marché (*Flourish Studio, Toucan Toco*) privilégient de nouveaux usages dans l'air du temps : personnalisation et interactivité, temps réel et accès mobile, travail collaboratif, utilisation de l'intelligence artificielle... Ces nouvelles formes de "data storytelling" ont pour ambition d'aider à la prise de décision sans maîtriser pour autant des logiciels complexes, à partir d'interfaces simples dans lesquels l'utilisateur intègre ses données.

Le tableau ci-dessous récapitule les différentes méthodes à adopter en fonction des corpus à analyser et des objectifs stratégiques :

La valeur ajoutée des professionnels de l'information dans le processus d'analyse

Les professionnels de l'information n'ont bien entendu pas vocation à se substituer aux experts d'un domaine. Mais leur rôle transversal les place au cœur d'un maillage d'expertises présentes dans plusieurs départements des entreprises et des organisations. La valeur ajoutée des professionnels de l'information consiste ici non seulement en la

maîtrise des techniques et outils mentionnés plus haut, mais aussi à une interprétation judicieuse et non biaisée des résultats de la veille.

En effet, "faire parler la donnée" via les techniques de fouille ou datavisualisations ne veut pas dire la faire mentir ! La confusion entre corrélation et causalité peut provoquer des interprétations erronées ou fantaisistes. L'analyse automatisée et l'intelligence artificielle peuvent comporter (de même que l'intuition humaine) des biais cognitifs susceptibles de déformer la perception et la compréhension des situations et des enjeux stratégiques.

L'analyse de l'information trouve ainsi de nombreux domaines d'application : la détection d'innovation et de coopérations en veille technologique, l'identification de communautés et d'influenceurs en veille marketing, l'analyse des performances d'une entreprise en veille concurrentielle, la surveillance des avis des consommateurs en veille e-reputation, et bien d'autres. Si les outils automatisés s'imposent désormais à l'heure de la surabondance de données hétérogènes, l'expertise du professionnel de la veille apporte aux décideurs, chercheurs ou autres acteurs stratégiques, une vision claire de situations complexes.

Véronique Mesguich

mesguichveronique@yahoo.fr

<https://www.linkedin.com/in/>

véronique-mesguich-030561142/

Janvier 2020