



Plus de 700 millions d'utilisateurs Facebook et près de 400 millions pour Twitter : le Web est devenu un immense terrain d'échange et de partage. La compréhension de ce phénomène représente un enjeu majeur pour les entreprises privées comme pour les acteurs publics.

E-réputation & analyse de l'opinion : mode d'emploi

Environ trois-quarts des Français sont des internautes, et parmi eux la moitié va, tous les jours, sur un réseau social. Deux-tiers des internautes considèrent qu'Internet a changé la manière dont ils s'informent sur les entreprises et les marques, et, parmi eux, trois-quarts consultent l'avis d'autres clients avant de choisir un produit. Ces chiffres montrent, s'il en était encore nécessaire, qu'Internet et les réseaux sociaux ont changé la donne et que pour les directeurs marketing rien ne sera plus comme avant. Feu de paille ou véritable révolution ? Pour le savoir, répondons à quatre questions :

E-réputation : alternative ou complément aux approches traditionnelles ?

Depuis plusieurs dizaines d'années, l'analyse et le suivi de l'opinion passent largement par deux types de méthodes, quantitatives ou qualitatives. Chez la plupart des professionnels, ces deux aspects sont très séparés, y compris en termes d'or-

** Président de Lingway*

ganisation interne. Les études quantitatives passent par des sondages. La méthode est la suivante : on définit les caractéristiques de la population que l'on veut étudier (CSP, âge, région...) ; on définit un échantillon représentatif de cette population, de quelques centaines à quelques milliers d'individus ; on leur pose des questions avec des réponses à choisir dans des listes fermées, ce qui peut introduire un biais important. Les instituts d'études complètent donc cette approche par des éléments plus qualitatifs : l'introduction de questions ouvertes dans les sondages et l'animation de groupes de discussions visant à recueillir un matériau plus qualitatif. Mais ces méthodes sont coûteuses, longues à mettre en place, et reposent sur de très petits échantillons.

L'approche e-réputation a l'avantage de permettre une observation "en temps réel", ce qui est hors de portée des méthodes traditionnelles.

La situation est radicalement différente dans une approche "e-réputation" : on observe des conversations spontanées entre consommateurs sur les réseaux sociaux et le web en général, ce qui évite le biais des questionnaires ; les volumes de conversations peuvent être très importants ; l'avantage est que l'on peut, potentiellement, faire des analyses qualitatives sur de grands volumes ; l'inconvénient est qu'on ne connaît généralement pas bien la population observée, en tout cas moins bien que sur un échantillon construit spécifiquement pour les besoins d'une étude.

Compte tenu des différences d'approches entre études traditionnelles et e-réputation, il paraît donc clair qu'elles ne sont pas concurrentes, mais plutôt complémentaires. On pourra renforcer les conclusions d'une étude par une observation des conversations sur Internet ou, à l'inverse, valider par une étude des observations faites par l'approche e-réputation. Mais en tout état de cause, l'approche e-réputation a l'avantage de permettre une observation "en temps réel", ce qui est hors de portée des méthodes traditionnelles. Cela permet des approches nouvelles de gestion de crises, de dialogues avec les clients et prospects, d'une manière générale de tout ce qui ressort de ce qu'on désigne aujourd'hui par "community management".

e-réputation : quels objectifs ?

L'objectif est d'apporter des réponses aux cinq questions suivantes :

- De qui parle-t-on ? Quelles marques, produits, personnes, sociétés ?
- De quels sujets ? Quelles thématiques ? Quelle actualité ? Quelles offres ?
- Où en parle-t-on ? Dans quels médias ? Quels réseaux sociaux ? Quels sites ? Quels blogs ? Quels forums ?
- Qui parle ? Quel internaute ? Quel blogueur ? Quel journaliste ?
- Quels jugements ? Quels avis ? Quelles recommandations ? Quelles tonalités ?

e-réputation : quelles méthodologies et quels outils ?

Le "sourcing". La première question qui se pose lors de la mise en place d'un projet de e-réputation est celle du "sourcing", c'est à dire la détermination des sources à surveiller. Pour schématiser, on peut distinguer deux types de situations selon qu'on connaît déjà ou non une liste de sources à surveiller.

Dans le premier cas, on connaît un "bouquet" de sources qui sont déjà qualifiées et jugées intéressantes par l'utilisateur : des comptes twitter, des pages facebook, des blogueurs, des forums, des sites de presse ou institutionnels, ... L'identification de ces sources provient souvent de projets antérieurs, dans le cadre de veilles ou d'études spécifiques à l'entreprise. Selon les cas, on aura identifié des sources en une ou plusieurs langues.

Dans le deuxième cas, faute d'un bouquet de sources intéressantes, on utilise les moteurs de recherche sur Internet et les réseaux pour les identifier. Le choix des mots-clés est alors critique : s'ils sont trop restrictifs, on risque du silence, c'est à dire que l'on va rater des sources intéressantes. S'ils sont trop génériques, on va avoir un bruit important, c'est-à-dire des documents ou des sources sans intérêt.

La bonne démarche est de viser plutôt large, de collecter plutôt trop de documents que pas assez et de disposer d'outils de filtrage suffisamment puissants pour se ramener, par effet d'entonnoir, aux sources et documents pertinents. Ces outils de filtrage seront d'autant plus efficaces qu'ils seront dotés d'analyseurs linguistiques capables d'identifier les thématiques et les tonalités de ces conversations suivies sur Internet. Bien sûr, dans la plupart des projets on va combiner les deux méthodes, bouquets et mots-clés. Dans la grande majorité des cas, le client connaît au moins quelques sources intéressantes qu'il faudra prendre en compte. Et la méthode des mots-clés permettra de mettre à jour, au fil de l'eau, la liste des sources à suivre.

Quoi qu'il en soit, il y a deux écueils à éviter dans cette phase de détermination des sources : d'une part, ne pas tomber dans le mythe de l'exhaustivité. L'internet est trop vaste et surtout trop changeant pour que quiconque puisse prétendre à l'exhaustivité ; d'autre part, et à l'inverse, ne pas tomber dans le piège de l'échantillon représentatif de sources : pour les mêmes raisons de changement permanent des frontières de l'Internet et des réseaux, où il se crée en permanence de nouveaux sites, forums, blogs qui peuvent prendre de l'importance du jour au lendemain.

Méthodes et outils pour l'analyse

Un premier constat est qu'il est difficile de se passer d'outils informatiques. Dans la plupart des cas, les volumes à observer seront trop importants pour être à la portée, dans des conditions économiques raisonnables, d'analyses purement "manuelles". D'un autre côté, on ne peut pas non plus imaginer des solutions qui seraient à la fois totalement automatiques et entièrement fiables. L'état de l'art en TAL (Traitement Automatique des Langues), cette discipline qui fait appel à la fois à la linguistique et à l'informatique, ne peut pas garantir de tels résultats. En revanche, ces outils fonctionnent suffisamment bien pour filtrer et catégoriser ces données avec un niveau de qualité et de précision suffisant pour répondre à des besoins concrets des entreprises. Le niveau de qualité dépend de plusieurs facteurs : le niveau de langues employé sur les sources (il est plus facile d'analyser un blog bien écrit qu'un dialogue à l'emporte pièce dans un forum), le niveau de précision recherché dans la catégorisation, les langues suivies, etc. En général, on peut considérer qu'il varie autour de 70 à 80 % de qualité. A partir de là, deux options sont possibles. Soit on se contente de ce niveau de qualité, qui donne une première approche du corpus de documents et messages de toutes sortes qu'on a ramassés. Soit on a besoin d'une qualité certaine et une révision humaine est alors nécessaire. Le point important est que cette révision sera beaucoup plus rapide que si l'on avait tout fait "à la main".

Les outils d'analyse du contenu

Il ne peut être question dans le cadre de ce bref article de rentrer dans le détail des technologies de l'analyse linguistique. Deux exemples vont en montrer l'utilité et les grands principes. Nombreux sont les noms de marques qui sont aussi des noms communs (Orange), des noms de lieux (Kronenbourg, Orange en France, Vinci en Italie), des noms de personnes (Vinci, Chanel, Heineken). Certaines marques cumulent les ambiguïtés : AXA est une princesse espagnole, une tribu amérindienne, un mouvement politique espagnol, un personnage de bande dessinée, une expression qui revient dans de nombreuses expressions

mathématiques, le passé simple du verbe "axer" et sûrement d'autres choses encore... Un système de e-réputation doit être capable de lever ces ambiguïtés en fonction du contexte.

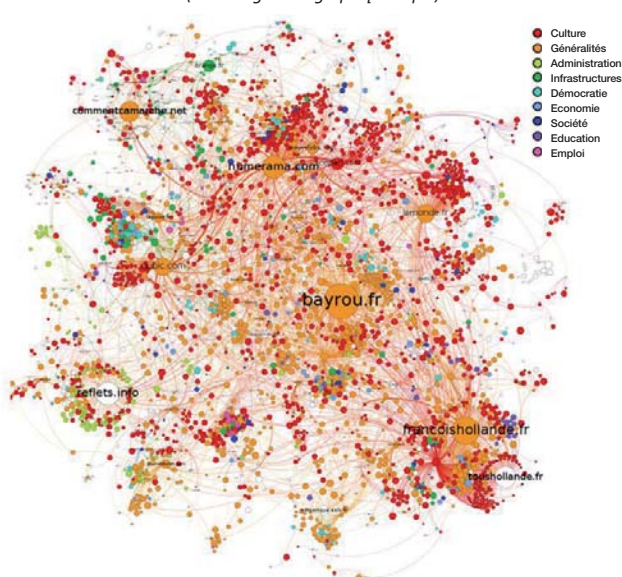
Un autre exemple porte sur l'analyse de tonalité : une approche naïve se baserait sur la seule tonalité des mots pour déterminer la tonalité d'un avis. Ainsi une phrase comme *"ce somnifère me donne des angoisses"* serait négative parce que le mot "angoisse" est noté comme tel dans un dictionnaire. Mais des phrases comme *"ce somnifère calme mes angoisses"* ou *"depuis que je prends ce médicament mes angoisses ont disparu"* sont positives alors que *"ce somnifère ne calme plus mes angoisses"* est négatif... On voit bien, intuitivement, qu'une analyse linguistique un peu fine est nécessaire pour éviter des contre-sens.

L'identification des influenceurs

Si tout le monde peut s'exprimer, tous n'ont pas la même influence, le même impact sur l'opinion. Il est donc important de repérer les "influenceurs". Plusieurs types de méthodes sont généralement utilisés :

- des classements d'audience sont publiés par des instituts spécialisés, dans la lignée de ce qui est fait pour les médias traditionnels ;
- la plupart des réseaux sociaux (Twitter, Facebook, etc.) donnent accès à des données indiquant le nombre d'abonnés ("amis", "fans", etc.) au compte d'une personne physique ou morale ;
- une analyse des relations de citations entre les sources donne également une idée de la formation des communautés et groupes d'influence.

(Technologie cartographique Gephi)



La carte permet par exemple de visualiser comment des comptes Twitter s'organisent autour des sites de référence lors de la dernière campagne électorale, la couleur de chaque point, qui identifie un utilisateur Twitter, étant déterminée en fonction de ses thèmes de prédilection identifiés par l'analyse des tweets émis.

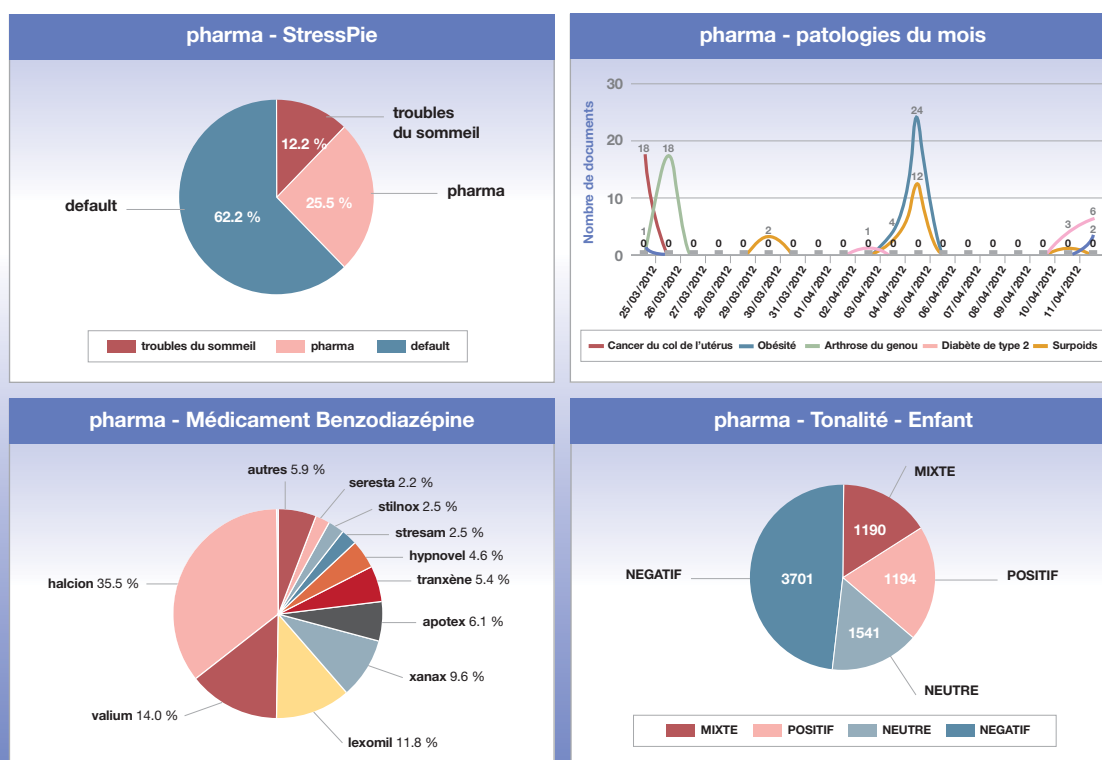
e-réputation : quels bénéfices pour une direction marketing ?

Nous n'en sommes qu'au début de l'ère des réseaux sociaux. La période d'il y a quelques années où nombreux, surtout les générations actives dans la vie professionnelle, émettaient des doutes sur leur utilité et leur pérennité, est derrière nous. Il est maintenant évident que leur influence, dans la plupart des domaines de la vie sociale va être de plus en plus importante. Pour l'entreprise, et notamment pour les directions marketing, ce sont de nouveaux enjeux qu'il n'est pas possible d'ignorer : mettre en place un suivi de son image ou de sa réputation sur Internet ; faire le même suivi sur ses concurrents, ses partenaires, ses fournisseurs ; mettre en place des systèmes d'alerte pour anticiper sur des crises naissantes ou potentielles. Mais les mêmes méthodes d'analyse s'appliquent aussi à d'autres champs que l'Internet : par exemple certaines entreprises ana-

lysent les messages professionnels circulant sur les réseaux sociaux d'entreprises pour identifier les idées nouvelles, les thèmes importants du moment, etc. Autre sujet, sans doute plus critique, est le suivi des messages de clients via les outils de GRC (Gestion de la Relation Client), où l'analyse des avis est primordiale.

Ces systèmes permettent de construire des baromètres, tableaux de bords et autres indicateurs qui vont permettre de faciliter les prises de décision. Car, au delà du recueil et de l'analyse de ce qui se dit sur Internet et les réseaux sociaux, c'est bien ce qui en découle en termes de décisions de gestion qui est important. Enfin, on conclura en remarquant que si la problématique de la e-réputation est née dans le domaine du B to C, le phénomène a pris de plus en plus d'importance, d'abord outre Atlantique, mais aussi en Europe dans le cadre du B to B, où les enjeux sont tout aussi importants. ■

**Lingway est l'éditeur de la plateforme de veille et d'analyse sémantique LINGWAY e-réputation, aujourd'hui utilisée quotidiennement par plusieurs grandes entreprises et organismes publics, et de grandes agences de marketing ou de communication.*



Exemple de tableau de bord réalisé avec LINGWAY* e-réputation