

(... sourcing ...)

Qualifier une source d'information et évaluer une information

indicateurs et outils

Le foisonnement actuel de l'information est indéniable. La théorie de la longue traîne, et dans son sillage la constellation de blogs, de réseaux sociaux, de forums, d'avis de consommateurs, donne lieu à la création massive de contenus générés par les internautes .



Cette création de contenus par les internautes est également dénommée "User Generated Content" ou UGC

Crédits photos :
<http://www.flickr.com/photos/sklathill/> sous licence Creative Commons : Attribution-Share Alike 2.0 Generic

Les problématiques de validation et de qualification de ces informations et des sources les diffusant est, plus que jamais, une préoccupation forte des veilleurs et des documentalistes. La littérature édicte 5 critères traditionnels d'évaluation de l'information

- L'exactitude
- L'autorité intellectuelle de l'auteur ou de l'éditeur
- L'objectivité
- L'actualité (information mise à jour)
- La couverture (sujets traités superficiellement ou en profondeur)

L'enjeu de cette évaluation est d'apporter une méta-information permettant de pondérer l'importance à accorder à une information avant que cette dernière ne soit prise en compte dans une décision. Ne pas se laisser désinformer, tout simplement.

Bien que cela puisse sembler être une évidence, de trop nombreux dispositifs de veille n'intègrent pas de qualification des sources ou des informations. Pourtant il existe des moyens, ou en tout cas des bons réflexes à avoir, afin, dans la mesure du possible, de répondre aux questions aussi essentielles que : cette source est-elle fiable ? Quelle est son autorité ? Cette information est-elle exacte ?

Information

*** (2) On se souviendra par exemple de l'Opération Volutes – Le Nouvel Observateur – 24 août 2005 – "Essec contre HEC" <http://tempsreel.nouvelobs.com/actualites/200508024.OBS7336/>**



*** (3) Parfois les données sont privées sur le site Afnic.fr, en particulier lorsqu'il s'agit de particuliers.**

Première étape

Commençons donc par les questions à se poser sur l'identification et la qualification des sources d'information, en l'occurrence d'un site Web.

La première de ces questions est : qui est le propriétaire ou l'auteur d'un site web, d'une information ?

Question incontournable et essentielle alors que les internautes se livrent à une course effrénée au personal branding ou au contraire usent de l'anonymat potentiellement offert par le web, dans des manœuvres parfois extrêmement agressives* (2).

Les qualités liées à l'éditeur d'un site web permettront souvent de considérer que l'information produite sur ce site hérite de toute ou partie de ces mêmes qualités. On aborde donc ici la façon d'établir un critère d'autorité intellectuelle de l'auteur dont le préalable nécessaire demeure son identification.

La première des choses à faire est donc de regarder sur le site en question si figurent des rubriques du type : « Qui sommes nous ? », « Mentions légales », ...

Lorsqu'il n'est fait nulle mention sur le site web de l'éditeur ou de l'auteur, le dernier recours reste l'utilisation d'un WHO IS. Tout possesseur d'un nom de domaine doit en effet déclarer son identité lors de l'achat ainsi qu'un certain nombre d'informations. Le registre de ces données est accessible grâce à l'interrogation des bases de données des différents organismes en charge de la gestion des noms de domaine, le découpage des zones de ces responsabilités se faisant par extension de nom de domaine. Par exemple, c'est l'AFNIC qui est en charge de la gestion des noms de domaine en .fr et l'accès à leur WHO IS est accessible à l'adresse : www.afnic.fr/outils/whois. * (3)

L'on peut ainsi aisément obtenir les noms et prénoms du propriétaire du nom de domaine et donc du site ainsi qu'une adresse.

Les pièges à éviter

- Le Whois est inefficace sur les sous domaines offrant des pages personnelles ou blogs hébergés en SaaS par exemple du type : xxx.blogspot.com ou yyy.free.fr
- Parfois l'on peut obtenir les coordonnées du prestataire de service ayant réalisé le site qui a fait l'acquisition du nom de domaine pour son client.
- Si vous ne connaissez pas l'organisme gestionnaire d'une certaine extension vous pouvez passer par des outils tels que www.who.is qui interrogent les différentes bases de données de noms de domaine.

Maintenant que vous avez identifié le propriétaire d'un nom de domaine il vous suffit de vous renseigner sur cette personne afin d'en savoir plus sur ses références, ses activités. Dans ce cas présent l'UGC est d'un grand recours car il n'est pas rare désormais que les personnes se dévoilent sur des réseaux professionnels tels que LinkedIn.com, Viadeo.com ou Xing.com, moins professionnels tels que Facebook ou MySpace, voire même sur leur blog personnel ou bien encore sur des sites de micro-blogging tels que Twitter ou Jaiku. Et bien évidemment, Google reste également un incontournable sur cette phase de qualification : tapez simplement le nom et prénom du possesseur du site web pour voir régulièrement apparaître des informations le concernant.

Inquiétant : En mars 2009, Arvind Narayanan and Vitaly Shmatikov, de l'université d'Austin au Texas, publient un article intitulé « De-anonymizing Social Networks », montrant une méthodologie et un ensemble de règles statistiques de calcul permettant l'identification de l'auteur d'une information sur un réseau social à partir des interactions du pseudo utilisée ou de la façon dont l'information se rediffuse sur les réseaux sociaux. Le taux d'erreur estimatif est de seulement 12 %.



* 4 • On pourra par exemple pour cela utiliser Google Scholar, qui bien que ne couvrant partiellement le champ des publications universitaires, reste un outil pratique dans ce cas précis. Exemple : http://scholar.google.fr/scholar?as_q=&num=10&btnG=Recherche+Scholar&as_sauthors=Arvind+Narayanan&hl=fr

* 5 • Que l'on peut obtenir par exemple sur Yahoo Site Explorer (ex <http://siteexplorer.search.yahoo.com/fr/search?p=http%3A%2F%2Fwww.veillemag.com&fr=sfp&bwm=i>)

* 6 • Le Page Rank d'un site web est un indicateur calculé par Google et entrant en ligne de compte dans le classement de pertinence effectué par le moteur. Il s'agit grossièrement de la somme du nombre de liens pointant vers une page divisé par le nombre de liens dans la page sur laquelle est ce lien. Le PR va de 1 à 10 est croit de façon logarithmique. L'article à l'origine du PageRank est accessible à l'adresse : <http://www-db.stanford.edu/~backrub/google.html>

* 7 • De nombreux outils permettent de connaître le PageRank d'un site dont www.prchecker.info et la Google Toolbar.

* 8 • Les données statistiques de consultation des sites Web sont souvent non disponibles sauf pour les principaux sites (par exemple en France via Mediamétrie-eStat.)

Toutefois, Google Analytics permet désormais aux webmasters de donner accès à leurs statistiques, mais la pratique n'est pas encore très répandue.

* 9 • Le retwitt correspond à une citation sur Twitter, une rediffusion. On peut obtenir ce classement par compte Twitter via : www.retweetrunk.com (par exemple).

« L'auteur fait-il référence sur la thématique ? »

Une fois ce travail préalable d'identification effectué on pourra donc répondre à la question suivante : « L'auteur fait-il référence sur la thématique ? ».

Afin de répondre à cette question on s'appuiera tant que faire se peut sur des éléments quantitatifs : « L'auteur fait-il référence sur la thématique ? »

- Nombre d'années en formation diplômante sur le sujet abordé
- Nombre d'années d'expérience professionnelle en relation avec le sujet
- Nombre d'articles ou ouvrages rédigés par l'auteur sur la thématique
- Nombre de citations de ses ouvrages (bibliométrie)* (4)
- PageRank *(6) du site web (PR de la page d'accueil)* (7)
- Nombre d'abonnés au flux RSS lorsque cette donnée est accessible via les API de Google

Exemple :

<https://feedburner.google.com/api/awareness/1.0/GetFeedData?uri=<ID du flux>>
<https://feedburner.google.com/api/awareness/1.0/GetFeedData?uri=Actuelligence>

- Nombre de visiteurs uniques *(8) sur le site en question (lorsque cette donnée est publique ou accessible via par exemple les statistiques de Google Analytics ou de Xiti)
- Nombre d'abonnés au flux Twitter d'un auteur ou plutôt classement en termes de retwits *(9).

Etablir une échelle progressive d'autorité

On pourra ainsi envisager d'établir une échelle progressive d'autorité en fonction d'un indicateur synthétique que l'on pourra construire sur ces éléments quantitatifs. Cet indicateur ne sera en aucun cas cardinal mais simplement ordinal. Il permettra ainsi de classer différents auteurs / sources d'information sur le critère de l'autorité sur une thématique.

Les technologies actuelles permettent avec quelques heures de développement informatique d'automatiser la mise à jour de certains de ces critères, dont le nombre de liens entrants, le PR d'une page, les abonnés Feedburner ou Twitter.

La mise en place de ces automatisations et l'historisation des ces indicateurs peut permettre également de voir émerger des leaders d'opinion en travaillant sur le delta de ces indicateurs.

En dehors même de cet indicateur la collecte de ces éléments quantitatifs permet d'avoir un éclairage relativement objectif sur l'autorité de la source en question en se démarquant des notions subjectives que l'on croise parfois sous l'intitulé « gourou » ou « leader d'opinion » ou bien encore « experts »

L'autre critère traditionnel incontournable est bien évidemment celui de l'exactitude.

Il s'agit là du point le plus critique et le plus délicat à évaluer surtout lorsqu'on ne dispose pas de compétences métiers pour juger de la véracité d'une information.

En dehors donc de se référer à un expert du sujet pour avoir sa validation sur le fond on pourra prendre en compte le processus de création de l'information.

On se renseignera donc afin de savoir si l'information a été validée par un comité éditorial ou par un comité scientifique ou plus généralement si elle a fait l'objet d'un processus de contrôle (vérification téléphonique, contrôle sur place, ...)

Autre démarche importante : essayer de recouper l'information.

Si la même information est diffusée par plusieurs sources cela augmente la probabilité qu'elle soit fiable. On veillera toutefois à faire attention que l'information émane bien de plusieurs sources différentes. Par exemple si une dépêche AFP ou Reuters est reprise par différents médias, une seule source est toutefois à son origine : la dépêche en question . La généralisation de l'information numérique et sa simplicité de rediffusion transforme parfois l'identification de la source d'origine en véritable jeu de piste.

En cas de doute sur l'exactitude d'une information cela se doit d'être précisé lors de toute

transmission de cette information afin d'assurer que les destinataires aient bien conscience du risque potentiel lié à l'exploitation de cette information. Tout dispositif de veille se doit de qualifier ses sources d'information. Ces métas données, et particulièrement celles liées à l'auteur / éditeur permettent un pilotage plus aisé du dispositif de veille en focalisant son attention et en accordant plus de crédit à des sources faisant autorité.

La qualification sur des bases quantitatives et objectives permet justement, dans le cas de doute sur la fiabilité d'une information, d'apporter un éclairage sur la confiance dont on peut créditer le média ou l'auteur la diffusant.

En aucun cas toutefois, un tel indicateur ou bien encore l'ensemble de ces éléments quantitatifs, ne doivent se substituer au contrôle de l'exactitude d'une information et particulièrement lorsque cette dernière est jugée comme critique.

FRÉDÉRIC MARTINET
WEBMASTER DU SITE [WWW.ACTULLIGENCE.COM](http://www.actulligence.com) ET DIRECTEUR
BUSINESS UNIT MARKET & COMPETITIVE INTELLIGENCE, SPOTTER.

<http://www.flickr.com/photos/svachalek/> sous licence
Creative Commons Attribution-Share Alike 2.0 Generic

