

**Lettre professionnelle
Recherche et Référencement**

Numéro 163

Octobre 2014



SARL RESEAU ABONDANCE - Olivier Andrieu - Consultant -

3 rue des Châteaux - 67140 Heiligenstein -

Tél : 03 88 08 83 26 - Fax : 03 88 08 83 27 -

E-mail : olivier@abondance.com - Web : www.abondance.com

Recherche et Référencement

Numéro 163 - Octobre 2014

-- Lettre d'actualité mensuelle sur la recherche d'information et le référencement de sites Web --

Au sommaire ce mois-ci :

- > [Comment obtenir des vignettes de ses vidéos dans les SERP de Google ?](#) (12 pages)
- > [Kevin Richard : "SEObserver est un outil qui va faciliter la vie des référenceurs"](#) (7 pages)
- > [Combiner les données Analytics et Webmaster Tools : Pourquoi ? Comment ?](#) (7 pages)
- > [Le « Knowledge Vault » de Google : évolution ou saut quantique ?](#) (8 pages)
- > [Référencement, vie privée et droit à l'oubli](#) (4 pages)

- > [Revue d'outils moteurs et SEO](#) (4 pages)
- > [Revue d'URL](#) (7 pages)

Lettre parue le jeudi 16 octobre 2014

(prochaine parution : mardi 18 novembre 2014)

Le contenu de cette lettre est accessible sur la zone "Abonnés" du site Abondance, à l'adresse : <http://abonnes.abondance.com/>

La lettre "Recherche & Référencement" paraît aux alentours du 15 de chaque mois (un seul numéro pour les mois de juillet-août)

Pour tout renseignement : © Olivier Andrieu, olivier@abondance.com

IMPORTANT : Pour continuer de vous délivrer la meilleure information possible, la lettre "Recherche & Référencement" doit être utilisée au mieux par ses abonnés, qui doivent notamment se plier aux règles suivantes :

* Il est **interdit d'envoyer, par mail** ou autres systèmes de transmission, un exemplaire de la lettre à quelque format que ce soit, à une personne qui ne serait pas abonnée.

* Il est **interdit de diffuser à l'intérieur d'une entreprise, à plusieurs adresses e-mails ou sur un intranet**, le contenu de la lettre "Recherche & Référencement" si vous n'avez pas souscrit un **abonnement "Entreprise"** (<http://lettres.abondance.com/rechref.html>) qui est fait pour cela.

* Il est **interdit de transmettre vos login / mot de passe** à des personnes qui ne seraient pas abonnées. Pour information, les adresses IP des machines se connectant sont relevées et tout abus pourrait être sanctionné si ce phénomène ne pouvait être justifié par vos services techniques.

* Il est **interdit d'utiliser à des fins privées ou publiques tout contenu** de cette lettre sans notre autorisation expresse et écrite : envoi à des clients, des prospects, voire reprise intégrale du contenu en changeant le nom de la lettre et de la source (eh, oui, ça arrive...), etc.

De nombreux abus ont été constatés ces dernières semaines (merci à ceux qui nous les ont signalés) aussi, nous nous permettons de vous rappeler ces quelques règles de "savoir lire" :-)...

Merci pour votre compréhension et bonne lecture.

Toute l'information contenue de cette lettre est © Abondance / Olivier Andrieu. Toute reproduction ou distribution des informations, données et textes de cette lettre est interdite sans l'autorisation expresse du site Abondance. Toute reproduction ou représentation, intégrale ou partielle de ce contenu, faite sans le consentement de l'auteur, serait illicite. La loi n'autorise, que les copies ou reproductions réservées à l'usage privé du copiste et non destinées à l'utilisation collective, d'une part, et, d'autre part, que les analyses et les courtes citations dans un but d'exemple et d'illustration.

Sommaire du numéro 163 - Octobre 2014

Comment obtenir des vignettes de ses vidéos dans les SERP de Google ?	4
Kevin Richard : "SEObserver est un outil qui va faciliter la vie des référenceurs"	16
Combiner les données Analytics et Webmaster Tools : Pourquoi ? Comment ?	23
Le « Knowledge Vault » de Google : évolution ou saut quantique ?	30
Référencement, vie privée et droit à l'oubli	38
Revue d'Outils Moteurs et SEO	42
Revue d'URL	46

Comment obtenir des vignettes de ses vidéos dans les SERP de Google ?

[Retour au sommaire de la lettre](#)

Domaine :	Recherche	Référencement
Niveau :	Pour tous	Avancé

Les pages de résultats (SERP) de Google montrent (parfois, et moins qu'avant) des vignettes de vidéos permettant de donner une meilleure visibilité à un résultat déjà positionné. Mais les places sont de plus en plus chères et la nécessité de parfaitement optimiser les indications fournies au moteur pour vos fichiers vidéos est devenue primordiale. Voici donc un pas à pas qui vous explique notamment comment créer des fichiers Sitemap dédiés et intégrer des balises Schema.org spécifiques pour remettre en avant vos vidéos dans les SERP...

Obtenir un bon positionnement dans Google, c'est le principal enjeu des référenceurs. Mais derrière cet objectif, que l'on sait efficace, le référenceur prend aussi plus largement en compte la « visibilité ». Etre en tête, c'est bien, être visible, c'est mieux.

Ainsi, si gagner la guerre des positions permet d'atteindre le Graal, un site un peu moins bien classé peut tirer son épingle du jeu grâce à des batailles de tranchées où l'objectif est d'arriver à obtenir un affichage plus outrancier dans les résultats de Google.

Ainsi, le fait d'avoir un titre plus attrayant dans Google, une description mieux marketée, un fil d'ariane, un sitelink ou un « rich snippet » augmente généralement le taux de clic vers la page concernée.

Dans ce cadre, obtenir l'apparition d'une image illustrative est un *must*. Quoi de mieux pour attirer le regard qu'un beau visuel, surtout quand les concurrents n'en ont pas ? Va-t-on préférer être premier sur une requête, ou second, avec une jolie miniature ? Quel résultat sera le plus cliqué ?

La réponse n'est pas si évidente : elle dépend beaucoup de la nature des requêtes, de la qualité des visuels affichés, etc. Mais en tout cas, c'est un paramètre à ne pas oublier lorsque l'on souhaite aller plus loin que le simple classement, quand on pense visibilité avant tout. Car l'enjeu est avant tout de convertir, non ?

C'était le bon temps...

Ces derniers temps, on utilisait notamment l'authorship afin d'obtenir une image des auteurs de contenus dans Google. Mais fin Juin 2014, il y a donc seulement une poignée de semaines, cet affichage a été supprimé.


Il ne nous restait, à nous, référenceurs, presque rien pour faire la différence avec une photo miniature (*thumbnail*). En fait, il n'y avait même plus qu'une seule solution – hormis le fait de passer par Google Actualités - pour faire apparaître un visuel immédiatement à gauche d'un résultat de recherche. Pour cela, il fallait que les résultats correspondent à une page présentant une vidéo. En utilisant un sitemap spécifique à la vidéo, on parvenait sans trop de difficulté à afficher une belle imagerie, rendant de fait le résultat plus voyant, comme dans l'exemple ci-dessous.

Google bande annonce hunger games 3

Web Vidéos Actualités Images Shopping Plus Outils de recherche

Environ 2 580 000 résultats (0,35 secondes)

Actualités correspondant à **bande annonce hunger games 3**

 **Hunger Games 3 : Peeta Mellark, méconnaissable sur une ...**
June News - il y a 2 jours
D'ailleurs, 5 choses sont à retenir dans le trailer de Hunger Games 3. Tout d'abord, on apprend que Katniss Everdeen est devenue le leader ...

Plus d'actualités pour "**bande annonce hunger games 3**"

Hunger Games - La Révolte : Partie 1 - film 2014 - AlloCiné
www.allocine.fr/film/fichefilm_gen_cfilm=200088.html
Bandes-annonces et extraits Hunger Games - La Révolte : Partie 1. 01:37. Hunger Games - La Révolte : Partie 1 Bande-annonce VO. 72 345 vues. 01:40.
La Révolte : Partie 1 Teaser (3) - Casting - Bande-annonce Hunger ... - Photos

SNIPPET VIDÉO

Hunger Games - La Révolte : Partie 1 Bande-annonce (2) VF
www.allocine.fr/.../player_gen_cmedia=19548022&cfi...
16 sept. 2014
Bande-annonce Hunger Games - La Révolte : Partie 1 - Hunger Games - La Révolte : Partie 1, un film de ...

Exemple de résultats de recherche pour la requête « bande annonce hunger games 3 »

Mais c'était sans compter, depuis la fin Juillet 2014, la suppression par Google d'une très grande partie des visuels de vidéos dans les résultats « classiques » de recherche. Seuls les ressources issues de sites « reconnus » d'hébergement vidéo continuent à être affichées sous forme de photographies (avec un symbole de lecture et parfois la durée totale de la vidéo liée). Un grand coup de balai dans les SERP qui n'est pas vraiment plaisant pour tous ceux hébergeant de nombreux contenus vidéos...

Il faut dire que l'on pouvait assez facilement signaler à Google l'existence d'un contenu vidéo et profiter alors d'une meilleure exposition. Tout ce qui est contrôlable facilement par le webmaster finit toujours pas ne plus être pris en compte par Google. Tout est une question de temps...

Cependant, on gagne toujours à appliquer les « vieilles ficelles » favorisant la mise en avant des contenus vidéos. En effet, si l'internaute cherche une vidéo via l'onglet « vidéos » de Google (même si celui-ci est très certainement beaucoup moins utilisé que l'option « Web » par défaut), un visuel est bel et bien affiché. Pour tout le monde, certes, donc ce n'est pas à proprement parler un avantage... mais ne pas suivre les bonnes pratiques pourraient amener à disparaître tout simplement des ces pages de résultats vidéos. Qui voudrait prendre ce risque ? Et si jamais Google revenait en arrière ? On n'est pas à l'abri d'une inconsistance supplémentaire de la firme.

The screenshot shows a Google search interface with the query 'bande annonce lucy' in the search bar. The search results are displayed in the 'Vidéos' tab. The first result is from YouTube, titled 'LUCY Bande-annonce VF - YouTube', with a thumbnail showing a person in a dark setting and a duration of 2:24. The second result is from Allociné, titled 'Bande-annonce Lucy - Lucy Bande-annonce (2) VF - Allociné', with a thumbnail showing Scarlett Johansson and a duration of 2:23. The third result is also from Allociné, titled 'Bande-annonce Lucy - Lucy Bande-annonce VO - Allociné', with a thumbnail showing the word 'LUCY' and a duration of 2:29. The search results are dated from May 2014 to June 2014.

Exemple de résultats de recherche pour la requête « bande annonce lucy » en mode « vidéos »

Les vidéos issues de sites « d'autorités »

Depuis cet été, donc, on ne retrouve plus dans Google, en mode « classique » de recherche, les miniatures vidéo qui faisaient les beaux jours de beaucoup de sites de bandes annonces, de trailers de jeux vidéo et autres ressources animées.

Pourtant, les résultats de recherche ne sont pas dénués de tous affichages. Par exemple, on croise régulièrement des résultats mettant en évidence des ressources de Youtube, Dailymotion, Vimeo, Wat.tv, Orange, Ina.

Les sites les plus « évidents » de partages de vidéos profitent donc d'un effet d'aubaine, mais ce ne sont pas les seuls : une multitude d'autres sites reconnus qui mettent en ligne de façon régulière, voire systématique, des vidéos, profitent aussi de snippets vidéo. C'est par exemple le cas du site de recettes Marmiton ou encore d'Allocine, Koreus, jeuxvideo.com, qui sont aussi bien connus du public. Plus surprenant, certains fournisseurs plus confidentiels de bandes annonces de films qui ont la particularité de diffuser leurs contenus sur de nombreux sites de la thématique, profitent aussi de ce snippet particulier, par le biais de leurs affiliés.

The screenshot shows a Google search for "construire un bateau". The search bar contains the text "construire un bateau" and a magnifying glass icon. Below the search bar, there are tabs for "Web", "Vidéos", "Images", "Shopping", "Actualités", "Plus", and "Outils de recherche". The "Web" tab is selected. The search results show "Page 2 sur environ 2 310 000 résultats (0,26 secondes)". The first result is "Construire son bateau amorceur - Bateau amorceur d ..." from www.amorceur-carpe.com. The second result is "BATEAU - MATERNELLE - IUFM" from www.montpellier.iufm.fr. The third result is "[PDF] Construire un bateau qui flotte ressources - Académie d..." from www4.ac-nancy-metz.fr. The fourth result is a video snippet from YouTube titled "Tonnerres de Brest. Ils ont réussi à construire un bateau en ..." with a duration of 1:05. The fifth result is another video snippet from YouTube titled "[TUTO] Gmod - construire son propre bateau | Lutorial ..." with a duration of 4:10.

Exemple de résultats de recherche montrant des snippets vidéo pour Youtube sur la requête « construire un bateau »

The screenshot shows a Google search for "comment manger une mangue". The search bar contains the text "comment manger une mangue" and a magnifying glass icon. Below the search bar, there are tabs for "Web", "Vidéos", "Images", "Actualités", "Shopping", "Plus", and "Outils de recherche". The "Web" tab is selected. The search results show "Environ 274 000 résultats (0,28 secondes)". The first result is a video snippet from YouTube titled "Technique de cuisine : Préparer une mangue - YouTube" with a duration of 1:41. The second result is "Comment manger une mangue - wikiHow" from fr.wikihow.com. The third result is "Comment découper une mangue ? - aufeminin" from www.aufeminin.com. The fourth result is "Comment mange-t-on une mangue ? comment on l'ouvre - ..." from www.marmiton.org. The fifth result is a video snippet from Marmiton titled "Préparer une mangue en vidéo - Vidéo cuisine - Marmiton" with a duration of 4:46.

Exemple de résultats de recherche montrant des snippets vidéo pour Youtube et Marmiton sur la requête « comment manger une mangue »

The screenshot shows a Google search interface with the query "dernier concert des beatles". The search results are filtered by "Web". The first result is from "L'Internaute" with the title "Le dernier concert des Beatles - 29 août 1966". The second result is a video snippet from Dailymotion titled "The Beatles - Rooftop Concert (London Original Version ...)" with a thumbnail and a duration of 21:47. The third result is from Elpedro titled "Revivez le dernier concert des Beatles ! - Elpedro".

Exemple de résultats de recherche montrant des snippets vidéo pour la requête « dernier concert des beatles » pour Dailymotion

The screenshot shows a Google search interface with the query "clip musique". The search results are filtered by "Web". The first result is from Atlantico titled "Les 10 clips les plus chers de l'histoire de la musique ...". The second result is from Time Out Paris titled "Le sexe dans les clips de musique - Time Out Paris". The third result is from Koreus titled "Clip Gangnam Style sans la musique - Koreus" with a thumbnail and a duration of 3:51.

Exemple de résultats de recherche montrant des snippets vidéos pour la requête « clip musique » pour Koreus

Pour avoir le droit à un snippet vidéo, rien de mieux, donc, que d'être considéré comme un **site d'autorité**. C'est-à-dire, d'après ce que l'on peut constater empiriquement, un site reconnu du public, dont la mise en ligne « propre » de vidéos est pour ainsi dire assurée, et/ou de sites dont les vidéos sont largement diffusées sur le reste du web, et/ou qui mettent en ligne très régulièrement des contenus vidéos.

On a donc tout intérêt à rendre disponibles en partage nos vidéos, et si possible en très grand nombre. Cependant, ce n'est pas à la portée de tout le monde d'atteindre la légitimité d'un Allociné ou d'un Marmiton, si bien que l'on peut difficilement donner une marche à suivre pour atteindre un but un peu trop prospectif.

De plus, les snippets vidéo apparaissent avant tout sur des requêtes dont Google soupçonne qu'elles sont liées au besoin informationnel de l'internaute. Si l'on recherche « alien isolation » (un jeu vidéo récent), les résultats de recherche ne montrent pas de vignettes vidéos, tandis que « bande annonce alien isolation » met plusieurs pages en avant grâce à des miniatures vidéo.

Par contre, diffuser des vidéos issues des sites d'autorité permet de profiter de leur snippet vidéo dans le mode « vidéo » de Google. C'est un pis aller puisque l'on obtient un résultat sensiblement identique, que les vidéos soient issues de tels sites ou soient hébergées par nous mêmes.

L'idée de Google, qui suit celle de la suppression des images d'authorship, est vraisemblablement de proposer une expérience de recherche sur desktop plus similaire à celle que l'on trouve sur tablettes et smartphone, tout en préférant couper court à tout risque de proposer un contenu inadéquat en faisant le pari de retenir avant tout les sites de « confiance ». Certains pensent aussi que c'est un bon moyen d'améliorer le CTR des pubs Adsense, mais peut-être est-ce du mauvais esprit ?

Quoi qu'il en soit, Google continue à identifier les vidéos comme telles, que les résultats présentent ou non un snippet vidéo (un tour dans les résultats de recherche en mode « vidéo » est sans équivoque), si bien que l'on a tout intérêt à faire en sorte d'aider le moteur à reconnaître les contenus vidéos.

C'est également bien plus prudent car si la période est à la retenue au niveau de la mise en avant de visuels chez Google, la période actuelle est synonyme d'instabilité où tout peut changer d'un jour à l'autre sans préavis.

C'est bien parce que nous sommes référenceurs et que nous aimons utiliser bretelles et ceintures afin de nous assurer que tout sera bien pris en compte que l'on décide de respecter toutes les bonnes pratiques liées aux vidéos, car en réalité, dans la plupart des cas, les vidéos seront reconnues automatiquement par GoogleBot et seront rapidement visibles dans les résultats de recherche en mode « vidéos », sans rien faire. Mais qui peut le plus peut le moins. Alors...

Le sitemap vidéo

La première tâche technique que Google recommande de mettre en place pour s'assurer de la bonne prise en compte des vidéos postées sur un site, est de mettre en place un sitemap pour vidéos. Cela permet d'aider le moteur de recherche à identifier les différentes vidéos qui sont lisibles sur nos sites.

Google fournit bien sûr une explication détaillée pour mettre en place un tel sitemap. On la trouve, en français, à l'URL <https://support.google.com/webmasters/answer/80472?hl=fr>, mais cette page est actuellement boguée, et de fait non exploitable, il faudra donc de préférence se pencher sur la version anglaise pour obtenir l'information complète : <https://support.google.com/webmasters/answer/80472?hl=en>

Ceci dit, pas d'inquiétude, la mise en place d'un sitemap pour vidéo est plutôt aisée. Nous allons voir par l'exemple comment en construire un.

Imaginons que nous mettions en ligne sur notre site internet une vidéo montrant comment utiliser l'un de nos produits, le balais multi-brosses. Cette vidéo est lisible sur la page <http://www.example.org/balais.html>

Nous allons créer un fichier [sitemapvideo.xml](#) que nous plaçons à la racine du site.

Ce fichier propose le contenu suivant (le texte orange est l'explication de chaque point – à ne pas insérer dans le fichier ; seul le code en vert doit être présent).

```
<urlset xmlns="http://www.sitemaps.org/schemas/sitemap/0.9"
xmlns:video="http://www.google.com/schemas/sitemap-video/1.1">
```

Il s'agit de la déclaration du type de sitemap que nous créons. Ici, un sitemap vidéo.

```
<url>
```

Chaque bloc `<url>` représente une page du site. On peut ainsi insérer plusieurs pages dans un même fichier sitemap.

```
<loc>http://www.example.org/balais.html</loc>
```

L'url de la page concernée

```
<video:video>
```

Le bloc vidéo traite la vidéo présente sur la page concernée. On peut insérer plusieurs blocs `<video:video>` si la page propose plusieurs vidéos.

```
<video:thumbnail_loc>http://www.example.org/miniatures/1.jpg</video:thumbnail_loc>
```

Il s'agit de l'emplacement de la miniature (un fichier image), représentatif de la vidéo. Les formats souhaités sont ceux habituels pour le web : jpg, png et gif. On peut (de préférence), fournir de très grandes images, jusqu'à 1920x1080 pixels selon la documentation.

```
<video:title><![CDATA[Comment utiliser le balais multi-brosses ?]]</video:title>
```

Le titre de la vidéo. L'entourage CDATA permet de faire prendre en charge facilement les éventuelles balises html / entités html qui pourraient être présentes dans le titre. Google indique une taille maximale de 100 caractères.

```
<video:description><![CDATA[Voici en 10 étapes simples comment le balais multi-brosses s'utilise. Première étape : l'assemblage. 2e étape : etc.]]</video:description>
```

Il s'agit de la description de la vidéo. On peut indiquer jusqu'à 2048 caractères. On peut y placer un texte adéquat, ou au pire, utiliser une transcription textuelle de la vidéo pour aider le moteur à comprendre le sujet du média.

```
<video:content_loc>http://www.example.org/balais.mp4</video:content_loc>
```

L'URL de la vidéo. Si la vidéo est disponible en plusieurs formats, on peut ajouter un bloc `<video:content_loc>` par ressource.

A noter : Google autorise le cloaking sur la ressource (via reverse DNS par exemple). L'idée, non affichée dans la documentation, est de s'assurer que GoogleBot peut toujours lire la vidéo. Ce dernier point ne va en effet pas de soi car de nombreux sites interdisent la lecture des vidéos par des sites tiers.

Même si la logique voudrait que le sitemap ne prenne en charge que des fichiers présents sur un même nom de domaine, on n'hésitera pas à y insérer des ressources qui sont jouées sur vos pages, mais hébergées ailleurs (c'est une bonne manière de pouvoir pointer vers des vidéos que vous diffusez via un CDN, par exemple).

```
<video:player_loc allow_embed="yes" autoplay="ap=1">
http://www.example.org/player.flv?id=balais</video:player_loc>
```

L'emplacement du lecteur vidéo, tel que si l'on accède à l'url indiquée la vidéo se présentera en lecture. L'attribut `allow_embed` autorise Google à proposer une lecture directe à partir de ses pages de résultats (ne comptez pas dessus), et l'attribut `ap` est une abréviation pour « autoplay » (1 ou 0, activé ou non activé). Si l'on a déjà indiqué l'url de la ressource vidéo via `<video:content_loc>`, on peut s'abstenir de mettre cette balise dans le sitemap. Vous pouvez insérer l'url d'un lecteur externe, hébergé ailleurs (sous réserve que la ressource distance soit bien explorable par GoogleBot).

```
<video:duration>600</video:duration>
```

Durée de la vidéo en secondes.

`<video:expiration_date>2014-11-05T19:20:30+08:00</video:expiration_date>`
Date à laquelle la vidéo expire, où il faut que Google la supprime de son index. Les cas d'utilisation de cette balise sont extrêmement rares. On peut ne pas l'utiliser dans la très grande majorité des cas.

`<video:rating>3.4</video:rating>`
Note de la vidéo sur 5 points. Facultatif, on peut vraiment s'en passer.

`<video:view_count>5469999</video:view_count>`
Nombre de visionnages de la vidéo. Facultatif, on peut vraiment s'en passer.

`<video:publication_date>2010-12-05T19:20:30+01:00</video:publication_date>`
Date de mise en ligne de la vidéo, sous la forme AAAA-MM-JJ ou AAA-MM-JJThh:mm:ss+TZD ou TZD correspond à la timezone. Pour l'europe centrale, cette dernière valeur sera donc 01:00.

`<video:family_friendly>no</video:family_friendly>`
Si la vidéo doit être filtrée par le safeSearch de Google, on indique « no ». Sinon, la balise est facultative.

`<video:restriction relationship="allow">FR</video:restriction>`
Cette balise sert à indiquer dans quels pays la vidéo peut, ou ne peut pas, être diffusée. Si l'attribut relationship est égal à « allow », alors les pays indiqués seront considérés comme étant les seuls à pouvoir diffuser la vidéo. Si l'attribut relationship est égal à « deny », les pays indiqués seront considérés comme ne pouvant pas diffuser la vidéo. C'est une balise facultative (par défaut la diffusion est globalement autorisée). La liste des pays s'indique au format ISO 3166 (cf. http://en.wikipedia.org/wiki/ISO_3166), avec un espace entre chaque abréviation de pays.

`<video:tag>balais</video:tag>`
Un tag est un mot-clé décrivant la vidéo. On peut ajouter de multiples balises `<video:tag>` (jusqu'à 32). L'idée est d'aider Google à comprendre le contenu de la vidéo.

`<video:category>entretien</video:category>`
Cette balise permet de catégoriser la vidéo. On peut y insérer une chaîne de 256 caractères. L'idée est d'aider Google à classer la vidéo.

`<video:gallery_loc title="Tous les guides produits">http://example.org/galerie</video:gallery_loc>`
Parfois une vidéo fait partie d'une galerie de vidéos, on peut alors indiquer l'URL de cette galerie à l'intérieur de la balise, tandis qu'on peut en indiquer le titre via l'attribut title.

`<video:price currency="EUR" type="own" resolution="HD">9.99</video:price>`
Si la vidéo n'est pas accessible gratuitement et qu'il faut l'acheter ou la louer on peut utiliser cette balise. Le prix s'indique à l'intérieur de la balise, la monnaie concernée dans l'attribut « currency » au format ISO 4217 (cf. http://en.wikipedia.org/wiki/ISO_4217), le type d'achat dans l'attribut « type » (« rent » pour location et « own » pour l'achat ferme), la résolution dans l'attribut « resolution » (« HD ou « SD » : Haute Définition et Définition Standard). Si l'on a plusieurs prix de ventes (selon les monnaies) et/ou plusieurs types d'achats (location et achat ferme), on peut utiliser plusieurs fois la balise `<video:price_currency>`.

`<video:requires_subscription>yes</video:requires_subscription>` Indique s'il faut être membre pour pouvoir visionner la vidéo (« yes » ou « no »).

`<video:uploader info="http://www.example.org/myself">Moi</video:uploader>`

Ici, on indique le nom de la personne ayant mis en ligne la vidéo, et dans l'attribut info l'URL de profil de cette personne sur le site. Cette balise est facultative.

```
<video:live>no</video:live>
```

La vidéo est-elle transmise en direct ? (« yes » ou « no »)

```
<video:platform relationship="allow">web mobile tv</video:platform>
```

Sur quelles plateformes la diffusion de la vidéo est-elle autorisée ?

L'attribut « relationship » indique si les plateformes indiquées à l'intérieur de la balise doivent être autorisées ou non autorisées (« allow » pour autoriser, « deny » pour interdire). Les plateformes possibles sont web, mobile et tv.

```
</video:video>  
</url>  
</urlset>
```

Maintenant que le fichier sitemap est finalisé, nous utilisons les outils Google pour les webmasters (<https://www.google.com/webmasters/tools/home?hl=fr>) pour le soumettre au moteur de recherche (zone *Exploration* > *Sitemaps* > *Ajouter un sitemap*).

On voit donc ici que mettre en place un sitemap vidéo est aussi facile que de créer un sitemap classique de site. Ce serait dommage de s'en priver. Bien sûr, ce type de Sitemap peut être créé à la main si vous disposez de quelques vidéos sur votre site, mais il devra être automatisé dès que ce nombre croît avec le temps.

Les balises schema.org pour la vidéo, autour de chaque vidéo

Si créer un sitemap est un bon point de départ pour donner à Google la liste de toutes les vidéos d'un site, il s'agit d'une intervention dont le résultat est un fichier indépendant. Si l'on souhaite faire davantage d'optimisation on-page, on utilisera en complémentarité des marqueurs de vidéo sur chaque page, dès qu'une vidéo y apparaît.

Pour cela, on dispose des marqueurs de schema.org, et notamment du marqueur d'objet vidéo. Les spécifications de ce marqueur sont disponibles à l'adresse <http://www.schema.org/VideoObject>. Mais Google fournissant une interprétation légèrement différente des différentes propriétés du marqueur, nous allons donc plutôt nous concentrer sur le guide fourni par Google, lisible à l'adresse <https://support.google.com/webmasters/answer/2413309?hl=fr>. Et comme la version française de cette page est également boguée, nous nous tournerons plutôt vers son équivalente anglaise : <https://support.google.com/webmasters/answer/2413309?hl=en>.

Une fois que les marqueurs ont été implémentés, on peut vérifier qu'ils l'ont été correctement en utilisant l'outil de Google pour tester les données structurées, qui se trouve à l'adresse <http://www.google.com/webmasters/tools/richsnippets>.

Outil de test des données structurées

URL HTML

http://www.dailymotion.com/video/x275wyw_c-est-des-mains-c-est-l-episor

APERÇU Exemples

Sélectionnez l'onglet "HTML" pour afficher le code HTML récupéré et essayer de l'adapter.

Résultats de recherche Google Recherche personnalisée Google

Item

type: http://schema.org/videoobject

property:

isfamilyfriendly: true

url: http://www.dailymotion.com/video/x275wyw_c-est-des-mains-c-est-l-episode-01-ep01-s01_fun

thumbnailurl: http://s1.dmcn.net/HZYHF/1280x720-lmm.jpg

name: C'est des mains - C'est l'épisode 01 (EP01-S01)

description: Voici les 3 premiers sketches de l'univers des MAINrionnettes©. Suivez nous sur https://twitter.com/cestdesmains et

author: Item 2

playertype: Flash

width: 1000

height: 564

duration: PT3M12S

uploaddate: 2014-10-06T12:04:12+02:00

embedurl: http://www.dailymotion.com/swf/video/x275wyw

regionsallowed: AD,AE,AF,AG,AI,AL,AM,AO,AQ,AR,AS,AT,AU,AW,AX,AZ,BA,BB,BD,BE,BF,BG,BH,BI,BJ,BL,BM,BN,BO,BQ,BR,B:

Exemple d'affichage de l'outil de test des données structurées appliqué à une page de vidéo de Dailymotion.

Nous allons procéder par l'exemple pour comprendre comment mettre en place des données structurées liées aux vidéos. Prenons une page HTML dans laquelle une vidéo est présente, avec un code volontairement très épuré :

```
<html>
<head>
  <title>Introduction au SEO | Tutoriel vidéo</title>
</head>
<body>
  <h1>Introduction au SEO</h1>
  <video width="400" height="222">
    <source src="video.mp4" />
  </video>
  <p>Les bases du référencement web pour tous : pagerank, linking, optimisation on page, etc.</p>
</body>
</html>
```

Tout semble indiquer que cette page présente une vidéo, puisqu'elle contient une balise `<video>`, et l'on peut supposer que le H1 contient le titre de la vidéo.

Mais tout ceci n'est qu'une supposition, si bien que pour en être sûr, et donc s'assurer que Google comprend bien les choses, nous allons ajouter des données structurées pour signifier sémantiquement que nous avons ici affaire à une vidéo.

De la même façon qu'auparavant, le texte de couleur orange est présent à titre d'explication. Il n'est bien sûr pas présent dans le code final de cette page html.

```
<html>
<head>
```

```
<title>Introduction au SEO | Tutoriel vidéo</title>
</head>
```

```
<body>
```

```
<div itemprop="video" itemscope itemtype="http://schema.org/VideoObject">
On ouvre un bloc sémantique indiquant que l'on va traiter un objet de type VideoObject.
```

```
<h1 itemprop="name">Introduction au SEO</h1>
L'attribut itemprop a pour valeur « name ». On indique ainsi que la valeur à l'intérieur de la balise h1 est le nom de la vidéo.
```

```
<meta itemprop="duration" content="T1H10M34S" />
Cette balise meta est invisible pour l'internaute, mais elle fournit pourtant une information sémantique aux robots d'indexation. L'attribut itemprop a pour valeur « duration ». On s'intéresse donc à la durée de la vidéo. L'attribut « content » contient la durée exprimée au format ISO 8601 (cf. http://fr.wikipedia.org/wiki/ISO\_8601). Ici on indique une durée de 1 heure, 10 minutes et 34 secondes.
```

```
<meta itemprop="thumbnailUrl" content="miniature.jpg" />
L'attribut itemprop a pour valeur « thumbnailUrl ». L'attribut content désigne donc l'URL à laquelle se trouve une miniature représentative de la vidéo. On peut indiquer des résolutions élevées comme 1920x1080 pixels.
```

```
<meta itemprop="contentUrl" content="http://www.example.org/video.mp4" />
L'attribut itemprop a pour valeur « contentUrl », l'attribut content désigne donc l'URL à laquelle se trouve le fichier vidéo. Il ne faut pas hésiter à utiliser le balisage sémantique, même si les vidéos que vous avez placé sur vos pages sont hébergées chez un tiers (Dailymotion, Youtube, etc.).
```

```
<meta itemprop="embedUrl"
content="http://www.example.org/player.swf?video=video&format=mp4" />
L'attribut itemprop vaut « embedUrl », donc l'attribut content désigne l'url à laquelle on peut accéder à un lecteur diffusant la vidéo. C'est une balise facultative : dans notre exemple de page html, on utilise <video> pour diffuser la vidéo, il n'y a donc pas à proprement parler de lecteur, il n'y a donc pas de raison de l'utiliser..
```

```
<meta itemprop="uploadDate" content="2014-07-05T08:00:00+08:00" />
L'attribut itemprop a pour valeur « uploadDate », donc l'attribut « content » contient la date de mise en ligne de la vidéo, exprimée au format ISO 8601 (cf. http://en.wikipedia.org/wiki/ISO\_8601)
```

```
<meta itemprop="expires" content="2012-01-30T19:00:00+08:00" />
L'attribut itemprop vaut « expires », l'attribut « content » contient donc la date à laquelle la vidéo expire, c'est-à-dire lorsqu'elle n'a plus lieu d'être indexée. En pratique, c'est assez rare d'utiliser cette fonctionnalité d'expiration. Facultatif.
```

```
<video width="400" height="350">
<source src="video.mp4" />
</video>
```

```
<p itemprop="description">Les bases du référencement web pour tous : pagerank, linking, optimisation on page, etc.</p>
L'attribut itemprop est égal à « description », cela permet d'affirmer, sémantiquement, que le contenu de la balise p contient la description de la vidéo.
```

```
</div>
```

```
</body>
```

</html>

Voici donc une deuxième façon, après le Sitemap XML, de donner des indications à Google sur le fait que votre site propose des vidéos.

Conclusion

Mais obtenir des vignettes vidéos dans les résultats de recherche de Google n'est plus aussi facile qu'avant :

- Si vous êtes un acteur majeur du web, mettez en place un sitemap vidéo ainsi que les données structurées Schema.org adéquates. Si vous publiez régulièrement de nouvelles vidéos, on peut envisager que des vignettes apparaissent dans les résultats classiques du moteur de recherche.

- Sinon, vous avez tout intérêt à mettre en place le sitemap vidéo et les données structurées afin d'aider à la compréhension de vos contenus vidéos. Mais ne vous attendez pas à des miracles... Le snippet vidéo est dorénavant réservé aux sites envers lesquels Google a le plus de confiance. Dans certains cas rares, vous pourrez éventuellement obtenir un snippet vidéo en plaçant dans un player Flash la vidéo d'un tiers de confiance. Mais on peut parier que ce n'est qu'un effet de bord... qui ne saurait perdurer.

Les changements de ces derniers mois (suppression de la plupart des vignettes de vidéos, fin de l'affichage des vignettes d'authorship, mise en place d'un Knowledge Vault jalousement gardé, etc.) montrent que Google, plus que jamais, se méfie des optimisations faites par les webmasters et les référenceurs. Seuls les plus grands acteurs du web sont désignés comme dignes de confiance, au détriment des autres acteurs, plus petits, qui constituent pourtant l'immense majorité du volume d'information sur Internet. La « loi Google » devient de plus en plus restrictive et nous devons en tenir compte à l'avenir...

Guillaume Peyronnet est gérant de Nalrem Médias. Sylvain Peyronnet est co-fondateur et responsable des ix-labs, un laboratoire de recherche privé. Ensemble, ils font des formations, pour en savoir plus : <http://www.peyronnet.eu/blog/>

Kévin Richard : "SEObserver est un outil qui va faciliter la vie des référenceurs"

[Retour au sommaire de la lettre](#)

Domaine :	Recherche	Référencement
Niveau :	Pour tous	Avancé

Kévin Richard, bien connu dans le petit monde du SEO français, est en train de lancer un outil en ligne baptisé SEObserver, avec de grandes ambitions autour de l'audit de site web. Une bonne raison pour lui poser quelques questions sur cet outil, son fonctionnement et ses futurs projets...

Bonjour Kévin, peux-tu te présenter en quelques mots ?

Je suis Kévin Richard, petite trentaine, consultant SEO depuis 2007 environ et ex « black hat » qui a fini par laisser son chapeau dans la cour de récréation. J'ai travaillé dans l'e-réputation et également la sécurité informatique.

Je suis formateur SEO au sein du SEO High Level et désormais, je me lance à 110% dans le projet SEObserver (<http://www.seobserver.com/>), un outil qui ambitionne de devenir la tour de contrôle SEO fait par et pour les SEO, sur lequel je travaille depuis environ 1 an.



Pourquoi avoir créé SEObserver ?

Tout s'est fait très vite et a été dicté par un besoin : celui de comprendre comment fonctionne Google. A l'époque, je monitorais comme tout le monde les positions de mes sites web, mais ce travail avait un gros défaut : induire éventuellement de mauvaises analyses et donc de mauvaises décisions. Un exemple très concret : on peut se réjouir en se levant le matin en constatant qu'on a gagné une position, alors que la seule différence par rapport à la veille est que le site devant moi a disparu, et j'ai donc mécaniquement gagné une place alors que dans les faits je n'ai aucun mérite. Avec le suivi classique de position, c'est impossible à deviner.

Du coup je me suis dit qu'il fallait que je monitore également mes concurrents afin de voir comment ça se passe pour eux. Et lorsque je l'ai fait, j'ai découvert des choses insoupçonnées auxquelles je ne m'attendais absolument pas. Cela a considérablement élargi ma vision et ma compréhension du fonctionnement de Google, et surtout, cela m'a permis de voir débarquer des sites longtemps à l'avance alors qu'ils étaient toujours en page 3 ou 4.

Ensuite, je me suis dit que c'était trop dommage de ne m'intéresser qu'à mes mots-clés : comme dit Confucius, "L'homme sage apprend de ses erreurs, l'homme plus sage apprend des erreurs des autres" et j'ai décidé d'élargir à des mots-clés que je ne surveille pas spontanément, mais sur lesquels il y a beaucoup de mouvement : la plomberie, le casino, le porno, etc. Aujourd'hui, on se regarde tous un peu le nombril pour surveiller ses propres petites positions, mais avoir une vision plus large est absolument nécessaire.

Je me suis retrouvé à monitorer environ 3 000 mots-clés. Puis j'ai poussé la barre à 5 000 puis 7 500, mais je n'étais toujours pas satisfait car je passais à côté de nombreux mots-clés faiblement recherchés. Puis, lorsque j'ai réussi à obtenir un extrait des mots-clés les plus tapés sur le web d'après Adwords, un fichier de plus d'1,2 millions de mots-clés, j'ai décidé de surveiller les SERPs d'une « shortlist » de 50 000 expressions.

Cela m'a donné un niveau de données assez impressionnant et m'a permis d'être mieux informé que mes propres amis sur les positions de leurs sites. Pour l'anecdote, j'ai découvert avant certains consultants que leurs propres sites se positionnaient sur certains mots-clés.

Ensuite, surveiller les rankings, c'est bien mais reste assez limité. Cela donne une bonne base de départ pour commencer à investiguer sur un site, mais il faut continuer à tout faire manuellement derrière : récupération des backlinks, analyse du « passé » du site internet, etc. Je me suis donc dit « ok, j'ai les rankings, mais j'aimerais maintenant intégrer les données me permettant de comprendre ce qu'a fait tel ou tel site pour en arriver là ». Or, la bonne vieille pyramide du SEO nous indique qu'il faut s'intéresser à trois choses pour essayer de cerner les facteurs de référencement : l'architecture, le contenu et les backlinks.

J'ai alors décidé d'inclure ces données dans mon outil afin d'avoir tout cela à disposition. Ainsi, lorsqu'on suit de près un site web, on peut savoir s'il a touché à son architecture (suivi du nombre de pages indexées), à son contenu (suivi du nombre de mots, changements de balise Title et archivage de versions de son site pour analyse manuelle) ou à son linking externe.

Pour les backlinks, je cherchais la meilleure donnée possible, la plus complète et la plus qualitative et j'ai choisi celui qui est à mes yeux le meilleur fournisseur de backlinks sur le marché mondial (mais aussi un des plus chers), à savoir Majestic SEO. Enfin j'ai décidé d'enrichir les données de Majestic qui me semblaient pas assez exhaustives en rajoutant des données de PR, rang Alexa, liens sortants, etc.

Détail de taille : contrairement à 98% des outils sur le marché, j'ai l'autorisation d'inclure les données Majestic dans SEObserver, et de les donner à mon client final, sans que ce dernier n'ait besoin d'avoir un abonnement chez eux. J'ai bataillé longuement (et en anglais, c'est dire !) mais un accord a finalement été trouvé. Ca ne me semblait pas logique de devoir « brancher » un compte existant à l'outil, c'est comme si Total disait à ses clients « allez chercher du pétrole en Irak et puis ramenez-le nous, nous vous le raffinerons et le transformerons en essence ». A l'inverse, je préfère que les clients n'aient pas ce souci et qu'ils aient une offre clé en main.

SEObserver vise donc clairement à prendre la place des autres fournisseurs de backlinks, en y ajoutant des données difficiles voire impossibles à trouver ailleurs.

Quelles sont les principales fonctionnalités de cet outil ? Peut-être en raisonnant sur un exemple ?

En une phrase : SEObserver donne accès aux rankings de tout le monde sur chaque mot-clé, et décortique la stratégie mise en place pour en arriver là.

Il existe donc trois grandes fonctionnalités à l'heure actuelle, et je vais les détailler en prenant un exemple.

1) Une SERP machine, la machine à remonter le temps dans les SERPs : cet écran permet de voir l'évolution des SERPs entre deux dates pour un mot-clé donné. C'est le cœur de l'outil, son réacteur nucléaire et l'aspect le plus intéressant : on peut débarquer sur une niche en analysant son historique, alors même qu'on n'en avait jamais entendu parler auparavant !

Pour l'exemple, je veux m'intéresser à un mot-clé lié aux devis. Je tape au hasard « devis » sur l'outil, et une série de mots-clés me sont proposés, triés par Volume x CPC x Concurrence Adwords. Choisissons « devis travaux » :

devis assurance auto	274,126	22,200	12.6	0.98
devis mutuelle	47,400	2,900	16.51	0.99
devis assurance habitation	37,294	2,900	12.99	0.99
devis travaux	33,564	6,600	5.41	0.94
devis demenagement	29,791	4,400	6.98	0.97
devis assurance	29,546	3,600	8.29	0.99
devis assurance auto en ligne	26,701	1,900	14.34	0.98
devis	21,285	9,900	2.5	0.86
devis assurance voiture	20,560	1,600	12.98	0.99
devis auto	16,120	1,900	8.57	0.99

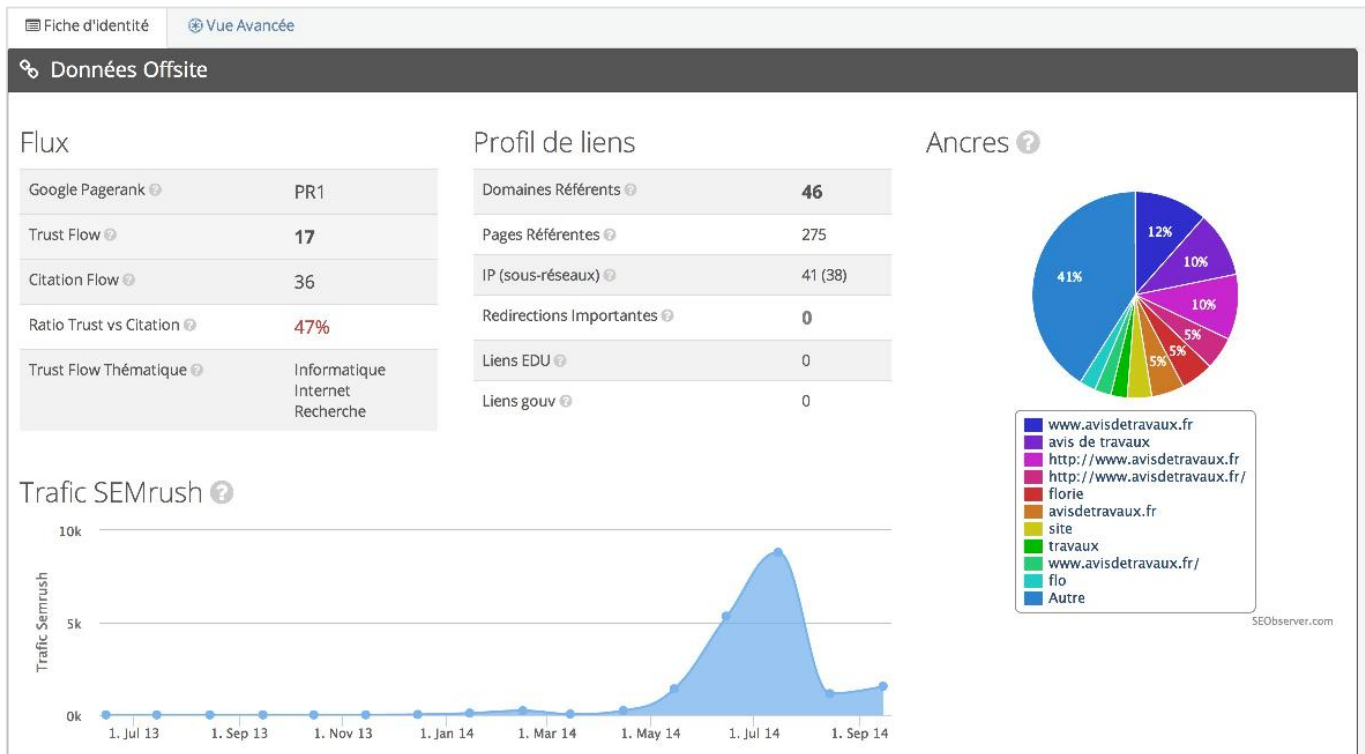
J'atterris alors sur la SERP Machine, véritable machine à remonter dans le temps, me présentant les SERPs d'aujourd'hui et d'il y a une semaine. Je change d'échelle pour avoir un maximum de recul et pour avoir tout l'historique. J'aperçois un site dans la colonne de droite qui n'était pas là il y a 7 mois puisqu'il est indiqué en « in ».

The screenshot displays three panels from the SERP Machine tool. The left panel shows search results for 'devis travaux' on 2014-03-21 (7 months ago). The right panel shows results for the same query on 2014-10-09 (18 hours ago). A central window, titled '202 days later', shows a LinkedIn logo and a line graph with a sharp spike, indicating a significant change in search results. In the right panel, the 6th result is highlighted in yellow and marked with 'in', indicating it was not present in the earlier snapshot.

#	Url	Δ
1	http://www.enchantier.com/	=
2	http://www.contactartisan.com/	=
3	http://www.123devis.com/	-5
4	http://www.easy-devis.fr/	-1
5	http://www.idealdevis.com/	-5
6	http://www.keltravo.com/devis-travaux/	-8
7	http://www.12travaux.fr/	-6
8	http://www.lillico-travaux.com/	-1
9	http://www.devis-travaux-en-ligne-ronchamp.com/	+5
10	http://www.devis-plus.com/	-2
11	http://www.kiloutou.fr/devis-travaux	+4
12	http://www.travaux.com/	+9

#	Url	Δ
1	http://www.enchantier.com/	=
2	http://www.contactartisan.com/	=
3	http://www.travaux.com/devis	+9
4	http://www.devis-travaux-en-ligne-ronchamp.com/	+5
5	http://www.easy-devis.fr/	-1
6	http://www.avisdetravaux.fr/	in
7	http://www.kiloutou.fr/devis-travaux	+4
8	http://www.123devis.com/	-5
9	http://www.lillico-travaux.com/	-1
10	http://www.idealdevis.com/	-5
11	http://www.simpledevis.fr/	+4
12	http://www.devis-plus.com/	-2

2) Je décide d'aller sur sa fiche d'identité avant de lancer une analyse plus importante, pour vérifier que ce n'est pas un site issu d'une redirection et pour regarder son « pedigree ». La fiche d'identité d'un site, c'est un peu le meilleur de Majestic, de Ahrefs et de Semrush en y ajoutant de la donnée issue des bases de l'outil.



Pas de redirections en vue (très très important !), et un nombre de domaines uniques assez cohérent (46 domaines). On trouve un ratio Trust Flow vs Citation Flow assez faible, et une thématique des backlinks (grâce au topical trust flow) assez éloignée des « devis travaux » on peut donc déduire que le linking n'a pas été forcément qualitatif. En même temps, les ancres nous indiquent que le linking a été fait de manière à éviter Penguin (pas de suroptimisation, c'est propre). Pour l'instant j'ai passé environ 1 minute 30 sur SEOobserver.

3) J'affiche la courbe positions vs backlinks afin d'identifier quels sont les backlinks qui ont eu le plus d'impact. On essaie toujours de qualifier les backlinks à travers des métriques telles que le Pagerank, le trust flow, etc. Mais au final, le meilleur backlink, c'est celui qui a un impact sur les rankings, donc comparer les calendriers de rankings et de linkbuilding est la meilleure approche pour détecter les backlinks « qui ont de l'effet » (même si pas forcément qualitatifs).



On voit sur le graphique ci-dessus les barres noires en bas qui représentent les nouveaux backlinks créés. On constate un passage de la 7ème à la 2ème page aux alentours du 13 juillet 2014. Je décide donc de resserrer la vue autour de cette date-là.

4) En dessous, j'ai déjà la liste des backlinks qui a été également filtrée entre le 1er janvier et le 13 juillet 2014. Je décide de faire abstraction des backlinks nofollow (décision contestable pour certains), et je me retrouve avec une shortlist assez surprenante, à laquelle je ne me serais jamais attendu.

Filtres actifs

Url is 51 Remettre à zéro

Advanced Filters >

<input type="checkbox"/>	URL	PR	Ancre	Flag	Nature	Cible Url	IP	Trouvé le
<input type="checkbox"/>	www.dmoz.fr/maison/renovez-et-optimis...	4	3 devis comparatifs d'artisans			www.avisdetravaux.fr/	162.159.247.31	2014-06-29✓
<input type="checkbox"/>	www.annuaire.com/portails-internet/av...	5	avisdetravaux.fr			www.avisdetravaux.fr/	195.60.188.51	2014-07-03✓
<input type="checkbox"/>	www.infos-vie-pratique.com/2013/07/02...	5	devis artisans			www.avisdetravaux.fr/	213.186.33.4	2014-06-12?
<input type="checkbox"/>	www.vivons-nature.com/nos-partenaires	3	devis artisans gratuits			www.avisdetravaux.fr/		2014-07-10✓

Résultat des courses : 4 backlinks, que je regarde manuellement et que trouve-t-on ? Du communiqué de presse SEO et de l'annuaire (attention, dmoz.fr n'est pas dmoz.org), et surtout, aucune ancre exacte « devis travaux » mais plutôt des ancres « devis artisans ». D'ailleurs, si on veut être sûr que ce sont bien ces backlinks et pas d'autres qui ont eu le coup de boost initial, il suffit de vérifier si le site se positionne sur ces mots-clés. Je tape « devis artisans » sur Google et ô surprise, notre ami se retrouve en 2ème place. Si je veux être plus précis je peux remonter en amont pour trouver ses backlinks antérieurs et analyser le « terreau » qui était en place au moment de la création de ces backlinks.

Terminé. Je connais en moins de 4 minutes 30 le cœur de la stratégie de ce site. Alors oui, ce n'est pas exhaustif, son SEO ne se résume pas à cela, et bien sûr, j'aurais pu regarder les versions archivées du site et ses changements « onsite », mais nous sommes ici dans un cas typique de ce que j'appelle du « spam doux », qui peut encore très bien marcher en 2014 (mais jusqu'à quand ?).

L'idée était de donner un rapide exemple de ce qui fonctionne encore à l'heure actuelle, sans pour autant prétendre être exhaustif aussi rapidement. Voilà ! Désormais, étant donné que j'ai épinglé ce site, j'aurai tous les matins la liste de ses derniers backlinks et je pourrai éventuellement m'en inspirer !

Ceci illustre une des possibilités d'utiliser SEOobserver. Il en existe bien entendu bien d'autres !

La communication autour de l'outil a été qu'il allait mettre "les équipes de la Search Quality Team de Google au chômage". Pourquoi cet angle d'attaque et pourquoi tant de haine ? :)) Cette vision est-elle appropriée ou est-ce juste du second degré ?

En fait, avec cet outil, mes attentes en termes de R&D SEO ont été tellement dépassées, j'ai tellement vu de choses auxquelles je ne m'attendais absolument pas (un peu comme dans l'exemple ci-dessus), que je me suis retrouvé avec une mine d'informations encore plus utile et bien plus pertinente que tout ce que peut avancer l'équipe qualité de Google, en tout cas d'un point de vue très très opérationnel.

Regarde avec l'exemple précédent, concrètement si tu poses la question suivante : « ok, que dois-je faire pour atterrir en première page sur devis travaux en moins de 6 mois ? ». La réponse de l'équipe qualité de Google sera : « Faites du contenu pertinent et utile pour l'utilisateur »... Bon, ok... :-)

La réponse de SEOobserver : « visiblement, un site avec peu de pages suffit. On peut compter sur environ une petite cinquantaine de backlinks, avec des ancres faiblement optimisées, splitté entre annuaires et communiqués de presse, en répartissant la campagne sur 6 mois et en ciblant quelques gros et beaux sites. D'ailleurs annuaire.com semble très bien fonctionner et une inscription dedans semble tout à fait légitime. Voilà la stratégie exacte qu'un site a utilisé récemment. »

Attention évidemment, le site a utilisé une technique risquée, et le fait de le mentionner ici même va sans doute déclencher une réaction des gens de chez Google qui vont le sanctionner manuellement, juste pour ne pas qu'on puisse dire « les CP et annuaires, ça marche toujours ! » alors que c'est évidemment le cas, si c'est bien fait.

Donc par rapport à ta question, je n'ai pas de haine envers eux. L'équipe qualité a un rôle d'évangélisation auprès des webmasters qui ont un niveau de SEO-awareness assez bas voire nul. Je ne compte pas les mettre au chômage bien entendu (même si eux mettent au chômage des entrepreneurs qui se font pénaliser), mais je souhaite que grâce à mon outil, les gens puissent enfin découvrir ce qui marche vraiment, loin des discours flous et dangereux qui peuvent conduire des sociétés droit dans le mur.

Où en est le développement de l'outil ? Est-il disponible officiellement et si non, quand le sera-t-il ?

Nous en sommes à la fin de la phase incroyablement chronophage mais nécessaire de corrections de bugs. L'outil marche déjà mais seulement si on lui indique correctement les données, et on veut le perfectionner pour qu'il gère tous les cas d'erreurs. Il devrait sortir d'ici fin octobre maximum.

Dans tous les cas, les premiers mois vont être très enrichissants en termes de nouvelles fonctionnalités, on va vraiment créer l'outil voulu par la communauté, et ce seront les clients qui décideront des nouvelles fonctionnalités à ajouter. Je suis moi-même consultant SEO, et j'ai posé la première pierre, maintenant c'est aux autres de me dire ce qu'ils veulent. J'ai en quelque sorte créé un cadre, et c'est aux autres de me dire ce qu'ils souhaitent y ajouter.

Quels sont les prix envisagés ?

Le premier prix est à 99€/mois et permet d'avoir accès à tous les rankings sur les 50 000 mots-clés et d'analyser 30 sites internet en analyse avancée (c'est-à-dire avoir le détail des backlinks comme dans l'exemple précédent), et un nombre illimité de sites en mode « fiche d'identité ». Enfin quand je dis illimité, c'est en usage « fair » et non automatisé évidemment.

Le deuxième prix est à 199€/mois et permet d'avoir accès au plan précédent et au listing complet de tous les mots-clés et en plus, d'analyser 100 sites internet en mode avancé. C'est le plan recommandé pour les agences. Nous allons rajouter la possibilité de générer des rapports en marque blanche dans les prochaines semaines.

Le troisième prix est à 499€/mois et permet encore beaucoup, beaucoup plus de choses, comme accéder à une pléthore d'outils cachés qui permettent de détecter des techniques black hat entre autres :)

L'idée, avec ces prix, est vraiment de mutualiser les coûts d'infrastructure (qui sont vraiment gigantesques) pour qu'on en profite tous ensemble.

Le SEO n'est pas une science exacte et des recettes qui fonctionnent sur un site peuvent ne pas marcher sur un autre. Comment un outil d'audit automatisé peut-il s'adapter à des cas de figures parfois très différents ?

SEObserver n'a pas la prétention de raisonner à la place d'un consultant, mais cherche plutôt à lui faciliter la vie. En gros, je ne cherche pas à remplacer un soldat par un robot, mais plutôt à armer ce même soldat avec un fusil d'assaut au lieu de son canif habituel.

SEObserver permet donc d'identifier des tendances sur les SERPs, et de faire remonter instantanément toutes les informations utiles sur les sites qui sont montés (ou sont descendus), pour les avoir à portée de main. Mon objectif est de faire en sorte de tout centraliser à un seul endroit pour ne plus avoir à faire des allers-retours ennuyeux entre différents outils. SEObserver a été conçu de manière à centraliser et concentrer le meilleur de tous les autres outils.

Autre question que j'aurais oublié ?

Ceux qui s'inscrivent sur la liste d'attente auront un gros avantage sur les autres, ils ne seront pas déçus :)

Merci Kévin, pour tes réponses !

Interview effectuée par Olivier Andrieu, éditeur du site Abondance <http://www.abondance.com/>.

Combiner les données Analytics et Webmaster Tools : Pourquoi ? Comment ?

[Retour au sommaire de la lettre](#)

Domaine :	Recherche	Référencement
Niveau :	Pour tous	Avancé

Google Analytics récupère pour nous de précieuses informations sur notre trafic. De plus, il est depuis quelques temps possible d'importer certaines données depuis le Centre pour webmasters de Google. Nous allons donc voir ici comment utiliser ces informations au quotidien dans une stratégie de référencement naturel, pour retrouver par exemple en partie les mots clés masqués dans le « Not Provided ». Mais pas que...

Préambule

Cet article se concentre sur l'association de deux outils appartenant à Google : sa solution de mesure d'audience Google Analytics (<http://www.google.fr/intl/fr/analytics/>), et son outil de gestion et monitoring de site Web, le Centre pour webmasters de Google (<http://www.google.fr/intl/fr/webmasters/>).

Cette association de données n'est en effet possible qu'entre ces deux outils car ils appartiennent à la même entreprise. Ceci est d'ailleurs également le cas avec les données de Google Adwords que l'on peut facilement associer dans un sens comme dans l'autre avec Google Analytics.

Les autres centres pour webmasters (Bing notamment) sont en effet indépendants par rapports aux solutions de Webanalytics, et on ne peut donc malheureusement pas faire de tels croisements de données.

Pourquoi associer ces deux outils ?

Le centre pour webmasters de Google a parfois mauvaise réputation en ne donnant qu'un accès restreint ou limité dans le temps aux données de son site. On peut donc tout naturellement se poser la question de l'intérêt d'associer ou non cet outil à Google Analytics.

En réalité, l'association des deux permet de faire ressortir certains éléments pour améliorer votre référencement naturel ou votre éventuel taux de conversion. Ce n'est d'ailleurs pas pour rien que le nouveau menu créé à cette occasion dans Google Analytics s'appelle « Optimisation du référencement ».

Voici les données que l'on peut obtenir par cette association :

- les expressions saisies par les internautes pour accéder au site ;
- le nombre de fois où chaque expression s'est affichée dans les pages de résultats de Google ;
- Le nombre et le taux de clics sur les pages positionnées dans Google ;
- La position moyenne des URL du site Web dans les résultats de recherche ;
- La liste des URL auxquelles les visiteurs ont accédé après avoir cliqué sur des résultats de recherche (les pages de destination).

Source : <https://support.google.com/analytics/answer/1308617?hl=fr>

Attention cependant : ceux qui utilisent souvent le Centre pour webmasters de Google vont se rendre compte qu'un grand nombre d'informations seront redondantes entre l'un et l'autre outil. Ne vous attendez donc pas à de nouvelles données auxquelles vous n'aviez pas déjà accès dans les "webmaster tools".

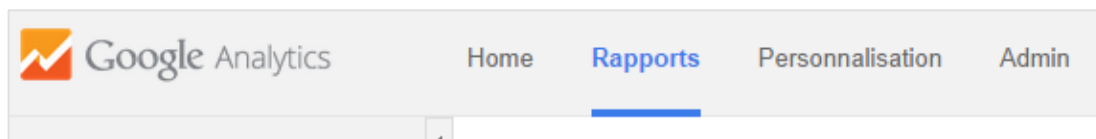
Au programme, on pourra notamment analyser les éléments suivants :

- récupérer une partie des mots clés cachés par le « Not Provided » ;
- Connaître et améliorer l'impact d'un changement de meta description ou de balise Title ;
- Pouvoir évaluer le positionnement moyen pour l'ensemble des utilisateurs pour les requêtes et pages déjà positionnées ;
- Connaître le taux de conversion de vos landings page en référencement.

Comment faire ?

La première chose à faire est de créer un compte Webmaster Tools et un compte Google Analytics pour votre site Web. Suivez sur chaque outil les procédures par défaut (très simples) pour activer correctement vos comptes.

Une fois cette étape réalisée, rendez-vous dans Google Analytics dans le menu d'administration (« Admin ») situé en haut de page, à droite.



Dans la colonne du milieu (« Propriété »), allez dans le menu « Paramètres du site ». Tout en bas de la page, vous aurez un espace « Paramètres des outils pour les webmasters ». Cliquez sur le bouton « Modifier » si vous voyez le texte « none » apparaître car cela voudra dire que vous n'avez associé aucun centre webmaster à votre profil Analytics.

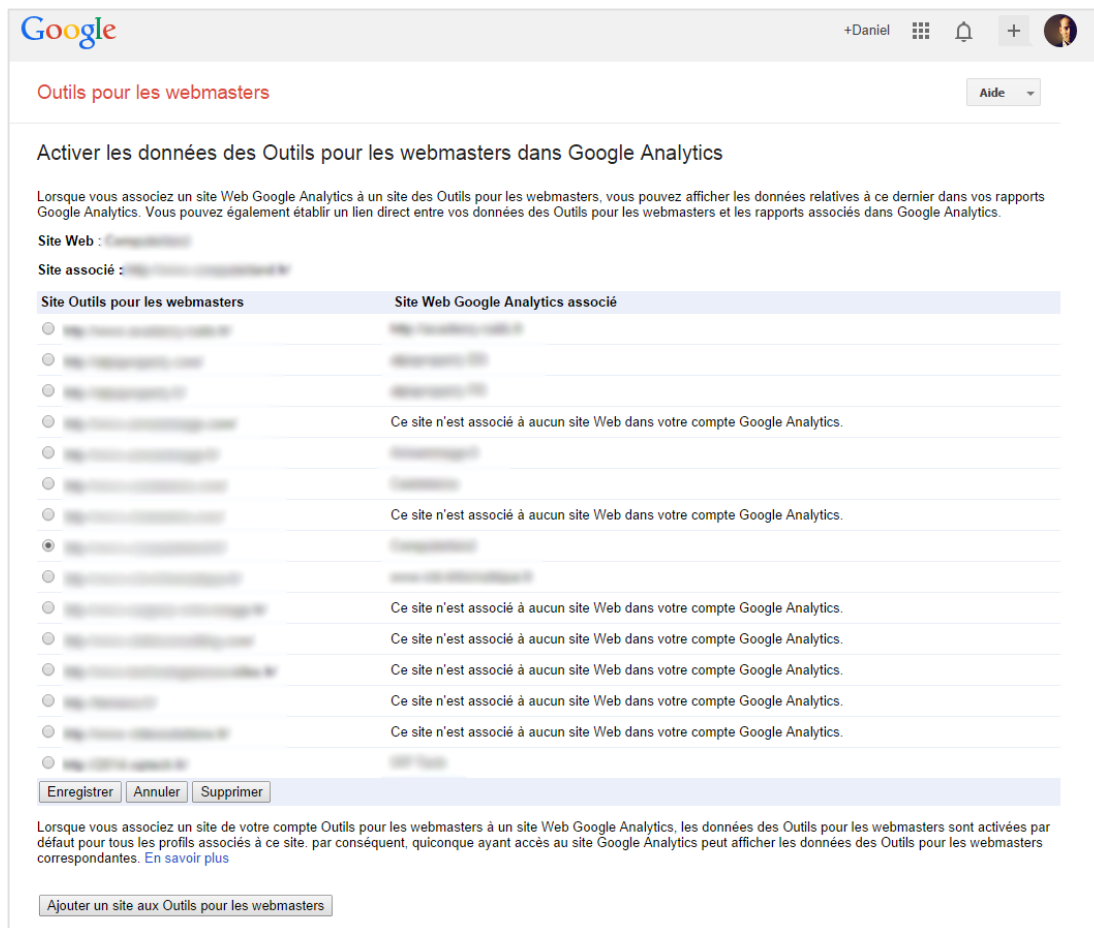
Paramètres des Outils pour les webmasters

Site Outils pour les webmasters facultatif ?

Si votre site a été validé avec les outils pour les webmasters, et que vous en êtes le propriétaire, vous pouvez associer ici les données des outils pour les webmasters. Vous serez ainsi en mesure de visualiser une partie de ces données dans certains rapports Google Analytics.

none [Modifier](#)

Dans la nouvelle page qui s'affiche, l'outil va afficher tous vos sites. Il vous suffira alors de sélectionner celui qui est pertinent pour l'associer à votre compte Google Analytics.



Vous verrez directement dans ce dernier menu les sites non associés avec le texte suivant dans la seconde colonne « Ce site n'est associé à aucun site Web dans votre compte Google Analytics ».

Remarque :

Les données qui s'afficheront dans les nouveaux menus dans Analytics auront les mêmes défauts que dans le centre pour webmasters :

- il faudra souvent attendre deux jours avant de voir les données s'afficher ;
- vous n'aurez accès aux données que sur les trois derniers mois. Vous ne verrez pas les données s'afficher pour l'année précédente et les années antérieures, hélas.

Sources :

- <https://support.google.com/analytics/answer/1308621?hl=fr>
- <https://support.google.com/analytics/answer/1308617?hl=fr>

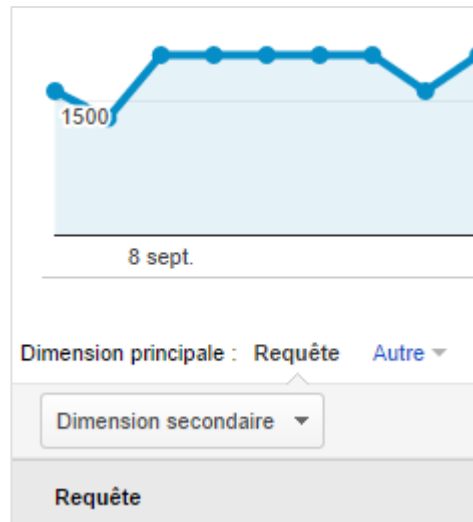
Quelle est l'utilité réelle de cette association ?

Les menus qui vont nous intéresser maintenant se situent dans la colonne de gauche, dans « Acquisition => Optimisation du référencement ». En réalité, les deux premiers sous-menus sont ceux qui seront réellement intéressants, et le troisième, dédié à la synthèse géographique, aura tendance à être secondaire. Voyons comment les utiliser :

Remarques préalables

Petite remarque avant de continuer : sachez que chaque menu affiche les données sous la forme d'un graphique, puis d'un tableau. Dans ce dernier, on peut appliquer des filtres de

manière dynamique. Ici, nous allons nous intéresser à l'un d'entre eux : la dimension secondaire.



Ces dimensions secondaires varient en fonction du menu où vous vous trouvez. Dans notre cas, elles sont peu nombreuses, mais l'une d'entre elles peut être intéressante : le « site Google ». Cela vous permet, pour tous les menus que nous allons présenter ensuite, de pouvoir afficher les données selon leur provenance. On pourra en effet filtrer les données du centre pour webmasters de Google en fonction de leur origine :

- Les recherches sur le Web en général ;
- Le moteur de recherche d'images ;
- Les recherches vidéo ;
- Les recherches pour mobiles.

Cela permettra donc d'obtenir un meilleur aperçu de son positionnement ou encore de son taux de clic en fonction du type de recherche naturelle qui aura été effectué.

Le menu Requêtes

Le premier menu est sans doute le plus intéressant. Il liste toutes les requêtes pour lesquelles votre site est apparu dans les moteurs de recherche (avec ou sans visiteur à la clé), avec à chaque fois :

- le nombre d'affichages (les impressions) ;
- le nombre de clics ;
- la position moyenne de ce mot clé ;
- le taux de clic.

Retrouver le Not Provided

Le premier intérêt est donc évident : en triant le tableau selon le nombre de clics, vous allez retrouver une très grande partie des mots clés cachés dans le « Not Provided », c'est-à-dire vos visites en provenance du référencement naturel.

Malheureusement, une fois votre liste de mots clés affichée, Analytics ne vous laissera pas la possibilité de connaître la page de destination associée, ce qui est un énorme handicap dans une optique de calcul du taux de conversion de votre trafic.

Connaître l'impact d'un changement

Quand on travaille son référencement naturel, on est parfois amené à changer la balise Title d'une page et sa balise meta description. Dans ce menu, vous pourrez en connaître l'impact de deux manières différentes :

- l'amélioration ou non de la position moyenne du mot clé ;
- l'amélioration ou non du taux de clic dans les résultats du moteur de recherche.

Evaluer le positionnement moyen

Tout le monde le sait désormais : Google personnalise les résultats en fonction de différents critères, notamment votre connexion à un compte Google, votre historique de recherche ou encore votre position géographique. C'est d'ailleurs pour cette raison que l'on constate fréquemment des différences de positionnement entre différents ordinateurs.

Il y a donc toujours une marge d'erreur entre le positionnement réel moyen, et le positionnement par défaut de Google. Ce menu vous donnera donc le réel positionnement sur l'ensemble des internautes ayant fait des recherches sur vos expressions clés.

Connaître les opportunités SEO

Un des éléments auquel on ne pense pas forcément dans ce menu concerne les opportunités SEO. C'est d'ailleurs ici le principal et meilleur atout de l'association des deux outils.

Dans nos logiciels de positionnement, nous avons une liste de mots clés à travailler, mais elle peut parfois être incomplète et on peut avoir oublié une expression clé.

Le menu « Requêtes » affiche en effet tous les mots clés qui sont présents dans les résultats de Google, que ce soit en 1^{ère} position ou en 20^{ème}. Il existe ainsi peut-être des mots clés très prometteurs situé en bas de première page, ou en seconde ou encore en troisième page, et qui pourraient pourtant attirer un trafic bien plus important.

Pour cela, on peut utiliser les filtres dynamiques avec la petite case de recherche située en haut à droite du tableau :



Cliquez sur « avancé », puis utilisez par exemple ces paramètres :

Inclure => Impressions => Supérieur à => 200

ET

Inclure => Position moyenne => Supérieur à => 4

ET

Inclure => Position moyenne => Inférieur à => 25

Dimension principale : Requête Autre ▾

Dimension secondaire ▾

Inclure ▾ Impressions ▾ Supérieur(e) à 200 ✕

et

Inclure ▾ Position moyenne ▾ Supérieur(e) à 4 ✕

et

Inclure ▾ Position moyenne ▾ Inférieur(e) à 25 ✕

+ Ajouter une dimension ou une statistique

Jouez sur les chiffres pour peaufiner votre analyse, mais sachez que ce type de filtre permet de faire ressortir immédiatement des mots clés déjà positionnés, qui génèrent beaucoup de demandes (d'impressions) mais qui ne sont pas encore dans le top 3.

On pourrait faire de même par exemple avec les mots clés positionnés dans le top 5, mais avec un faible taux de clic (donc avec un affichage pas suffisamment "sexy" pour l'internaute).

Le menu Pages de destination

Ce second menu présente toutes les pages de destination référencées et affichées dans Google, avec les mêmes données : impressions, clics, position moyenne et taux de clic. Impossible malheureusement d'associer la page à un mot clé précis.

L'intérêt de base est simple : vous allez savoir quelles sont les pages qui sont positionnées dans Google, à quelle position elles sont présentes et avec quel niveau de taux de clics.

Connaître l'impact d'un changement

On va également, dans ce cas, pouvoir mesurer l'impact d'un changement de balise Title ou meta description. Une fois le changement réalisé, vous verrez si le taux de clic et la position moyenne change ou pas.

Cela facilite aussi l'analyse de la qualité du résultat dans Google. Au-delà de la position d'un contenu dans Google, une bonne meta description peut en effet augmenter drastiquement le taux de clics et donc le nombre de visiteurs.

Savoir quels sont les termes qui génèrent du clic.

Enfin, dernier atout du menu Requêtes : pouvoir lister les termes utilisés dans vos balises Titles et meta Description pour savoir lesquels ont tendance à générer plus de clics.

Le concept est simple :

- listez toutes les pages de destination présentes dans ce menu (faites par exemple un export CSV avec le bouton d'exportation de données situé en haut de la page) ;
- dans Excel, listez pour chaque URL les balises Title et meta Description correspondantes en séparant chaque terme et en indiquant à côté le taux de clic correspondant ;

- faites alors un calcul pour savoir en moyenne quel est le taux de clic associé à chaque terme utilisé dans l'ensemble de vos balises.

Cette analyse est loin d'être fiable à 100%, mais peut parfois faire ressortir de réelles tendances. Dans son article en anglais, l'auteur Neil Patel a ainsi pu faire ressortir qu'en règle générale, ces termes augmentaient le taux de clic des pages positionnées :

- How to
- [List-related numbers]
- Free
- You
- Tips
- Blog post
- Why
- Best
- Tricks
- Great

Source : <http://www.quicksprout.com/2013/12/30/how-3-simple-google-analytics-reports-will-increase-your-search-engine-traffic/>

Le menu Synthèse géographique

Le dernier menu est peu utile. Il liste en effet le nombre d'impressions et de clics ainsi que le taux de clics par pays pour le référencement naturel.

On peut donc éventuellement découvrir des opportunités linguistiques, c'est-à-dire des pays dans lesquels nos résultats s'affichent et pour lesquels les contenus pourraient éventuellement être plus adaptés, avec par exemple :

- Une traduction si nécessaire ;
- Une ergonomie différente ;
- Des serveurs plus proches des visiteurs donc plus rapides ;
- ...

Conclusion

En nous permettant de lier et d'afficher directement dans Google Analytics les données du centre pour webmasters, Google nous facilite l'accès à certaines données.

En lisant cet article, certains d'entre vous se seront dit que ces données étaient déjà présentes dans les Webmaster Tools, et que cela n'apporte rien de plus. En soi, c'est vrai.

Mais l'interface plus simple et plus complète d'Analytics, notamment les filtres dynamiques que l'on peut appliquer beaucoup plus rapidement et facilement, joue en faveur de cette action. L'exemple donné ici pour trouver des opportunités SEO est en effet bien plus efficace et rapide que ce que l'on peut faire de base dans le centre pour webmasters.

Au final, il existe un seul avantage à consulter les données dans les Webmaster Tools plutôt que dans Google Analytics : dans le menu « Trafic de Recherche > Requêtes de recherche », puis dans l'onglet « Pages les plus Populaires », vous pourrez parfois voir quels mots clés sont associés à quelles pages de destination. Et ces données-là ne sont pas disponibles dans Analytics !

Daniel Roch, Consultant WordPress, Référencement et Webmarketing chez SeoMix
(<http://www.seomix.fr/>)

Le « Knowledge Vault » de Google : évolution ou saut quantique ?

[Retour au sommaire de la lettre](#)

Domaine :	Recherche	Référencement
Niveau :	Pour tous	Avancé

Le Knowledge Graph, basée sur l'outil Freebase, a été lancé il y a deux ans par Google et permet au moteur de recherche, par détection d'entités nommées, de fournir directement des réponses à certaines requêtes informationnelles. Mais un nouveau concept, imaginé pour combler les lacunes du Graph, est en train de se mettre en place dans les laboratoires de recherche de la firme de Mountain View. Ce Knowledge Vault n'en est encore qu'à ses balbutiements mais il n'en est pas moins prometteur...

Le 16 mai 2012, Google annonçait sur son blog le lancement d'une nouvelle fonctionnalité qui marquait un tournant dans l'histoire du moteur de recherche : le Knowledge Graph (<http://www.abondance.com/actualites/20120521-11478-knowledge-graph-google-officialise-son-moteur-semantique.html>). Ce « graphe des connaissances » était présenté comme une immense base de données de faits qui permettrait à Google d'afficher directement toute une série d'informations utiles sur sa page de résultats.



Deux ans plus tard, il semble que Google ait atteint les limites permises par les techniques utilisées par le Knowledge Graph. La base de faits couvrant une partie limitée des informations cherchées par les internautes, Google cherche donc à l'étendre en utilisant des techniques plus sophistiquées. Cette nouvelle approche, Google l'a baptisée « Knowledge Vault » (<http://www.abondance.com/actualites/20140902-14231-knowledge-vault-futur-recherche-google.html>), et elle risque de changer radicalement le fonctionnement du moteur sur de nombreuses requêtes et dans de nouveaux cas d'utilisation.

Mais commençons par rappeler ce qu'est le Knowledge Graph, et pourquoi il a vite atteint ses limites.

Le « Knowledge Graph » et la réponse instantanée sur les requêtes informationnelles

Traditionnellement, on sépare les requêtes des internautes en trois grands types :

- Les requêtes navigationnelles (trouver l'URL d'une page bien déterminée. Exemple un utilisateur qui cherche la page d'accueil du Monde.fr, ou un post particulier sur un blog).
- Les requêtes transactionnelles (télécharger un logiciel, acheter un produit...).
- Les requêtes informationnelles (rechercher une information sur le Web : ces requêtes sont les plus fréquentes).

Pendant longtemps, les moteurs de recherche ont été conçus avant tout pour un usage navigationnel : les moteurs proposaient une liste de résultats sous formes de liens bleus cliquables. Ceci est effectivement parfait pour une requête navigationnelle, mais si l'utilisateur cherche une information, il lui faut cliquer sur les liens et chercher lui-même dans les pages ainsi ouvertes si l'information recherchée y figure bien (ce qui n'arrive pas toujours).

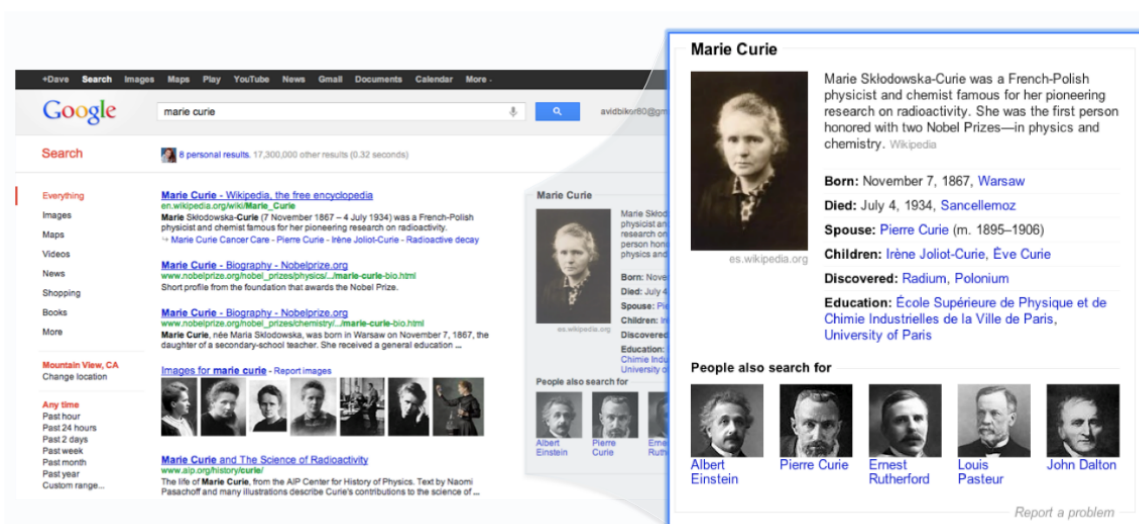
Entre 2004 et 2012, des tentatives plus ou moins réussies de création de « moteurs de réponse » ont vu le jour. Un « moteur de réponses », par opposition au modèle de moteur de recherche classique, donne directement la réponse à une question posée.

Mais pour que cela fonctionne, il faut que deux conditions soient remplies :

- Il faut comprendre la requête (la « question » de l'internaute) avant de pouvoir répondre.
- Il faut disposer de la réponse, ce qui suppose de disposer de bases de données particulièrement riches.

La première condition a nécessité des progrès dans la reconnaissance des entités nommées (les noms de lieux, de personnes, les raisons sociales de société). L'une des difficultés de l'exercice consiste à reconnaître l'entité figurant dans la requête quand sa désignation est ambiguë : quand une personne fait une requête contenant les termes « Taj Mahal », cherche-t-elle des infos sur un monument ou sur le musicien ? S'agit-il de Michael Jackson le chanteur ou son homonyme plombier à New York ?

La seconde condition, Google a réussi à la remplir en rachetant Freebase, un projet de base de données « Open » rassemblant une base d'informations sur les entités nommées d'une taille impressionnante.

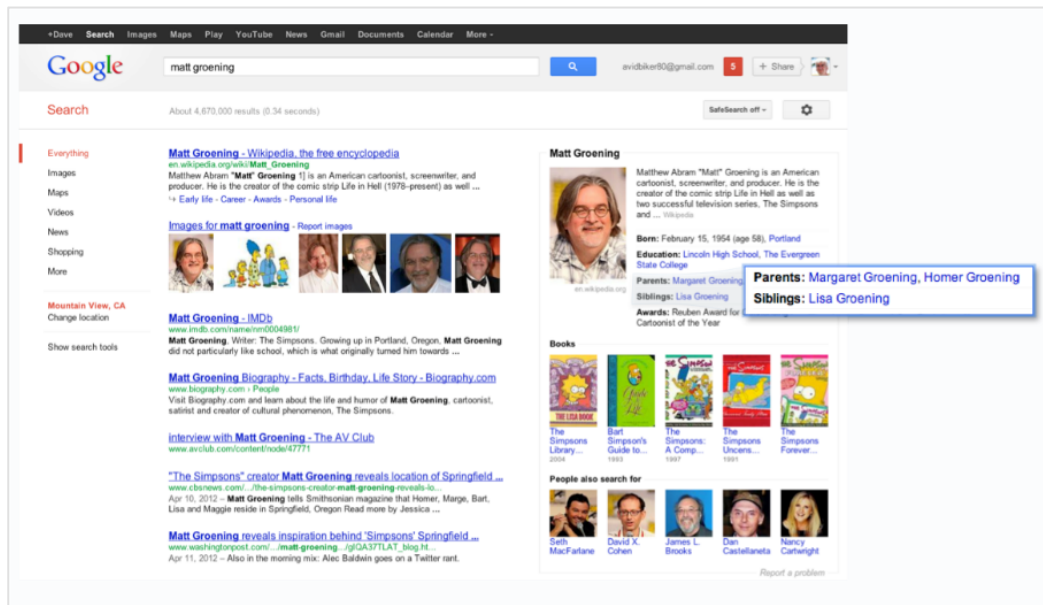


Exemple de données issues du projet Knowledge Graph : une requête sur "Marie Curie" fait apparaître une fiche complète sur la scientifique dans la partie droite de la page de Google, avec bio, dates de naissance et de décès, et de nombreux autres faits et informations.

Google a continué à développer Freebase à marche forcée, lui permettant de lancer le « Knowledge Graph » officiellement en mai 2012. Le Knowledge Graph est donc avant tout une extension de Freebase.

A fin 2013, la base du Knowledge Graph contenait :

- 500 millions d'entités (les nœuds du graphe) ;
- 3,5 milliards de faits (les arêtes du graphe) ;
- 1 500 types d'entités ;
- 35 000 types de relations.



Le Knowledge Graph stocke des entités, des faits à propos de ces entités, et des relations entre ces entités. Dans les informations sur Matt Groening, on trouve des faits (couple attribut <-> valeur : comme sa date de naissance) mais apparaissent également les relations entre Matt et ses parents (Matt Groening a pour parent Margaret Groening) .

Les limites de Freebase et donc du Knowledge Graph

Le cœur du Knowledge Graph est donc la base ouverte Freebase, qui est alimentée par un certain nombre d'acteurs du web. Malgré la diversité des sociétés et organismes qui l'alimentent (on y trouve aussi bien la CIA avec son célèbre « book of facts » que Netflix ou Wikipedia), des pans entiers de connaissance en sont absents.



Un aperçu des contributeurs les plus connus de la base Freebase.

Pour couronner le tout, les « fiches » y sont souvent incomplètes, ce qui limite beaucoup leur intérêt. En ce qui concerne les personnes célèbres par exemple, leur nationalité est non renseignée dans un cas sur quatre, leur lieu de naissance dans 29% des cas, et leur profession dans 32% des cas !

Relation	% unknown in Freebase
Profession	68%
Place of birth	71%
Nationality	75%
Education	91%
Spouse	92%
Parents	94%

*Un aperçu du taux de complétion des faits dans les fiches
Freebase pour les entités de type « célébrités ».*

Le taux de couverture finalement assez faible de Freebase limite considérablement les applications du Knowledge Graph. Les chercheurs de Google ont donc commencé à chercher des alternatives au crowdsourcing pour compléter leurs bases.

L'extraction des faits et des relations : un problème complexe

Depuis une dizaine d'années, des techniques ont été inventées pour extraire des « faits » à partir de l'analyse textuelle des pages web. Il faut avoir recours à une analyse syntaxique fine pour « désambiguïser » le contexte et identifier clairement les entités, les attributs associés aux entités et leur valeur. On peut aussi utiliser la hiérarchie DOM HTML (notamment la présence de tableaux ou de listes) et la mise en forme pour améliorer la détection. L'approche fonctionne de manière relativement satisfaisante.

Les techniques d'extraction d'information ont de plus fait des progrès dans plusieurs domaines. Au début, il fallait impérativement savoir ce que l'on cherchait pour le trouver : on ne cherchait que certains types d'entités, et que les valeurs de certains attributs liés à ces types d'entités. En réalité, la typologie des entités nommées n'est pas limitée aux types habituellement cherchés, les attributs sont très divers, et les relations extractibles également beaucoup plus nombreuses dans la pratique que celles que l'on pourrait imaginer *a priori*.

Les progrès ont concerné deux aspects qui permettent d'envisager une plus grande efficacité des méthodes d'extraction :

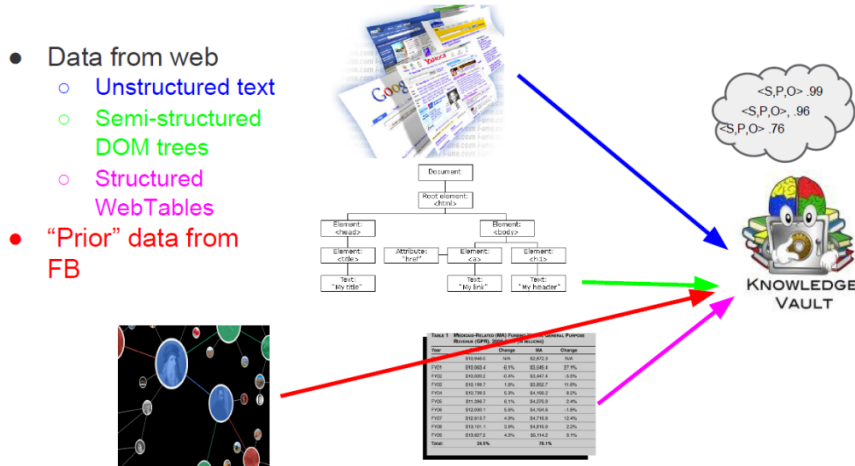
- **La détection automatique de candidats pour de nouveaux types d'entités**, d'attributs et de relations (notamment avec les méthodes d'OIE : *Open Information Extraction*). Les résultats sont loin d'être parfaits, mais cela augmente la couverture (le taux de succès) de l'extraction.
- **La détection automatique de motifs**, qui permet d'apprendre à extraire des informations sur plus de pages web sans qu'un humain ait besoin de paramétrer le système d'extraction pour chaque site web utilisé comme source d'infos.

Mais le problème est que les informations ainsi extraites ne sont pas toujours fiables, car toutes les pages web ne sont pas des sources dignes de confiance et exemptes d'erreur. Donc s'il est possible d'exploiter les données extraites du Web sur le papier, et d'accéder ainsi à un

volume de « faits » beaucoup plus conséquent, dans la pratique c'est le meilleur moyen pour stocker dans les bases de connaissance des données qui vont se révéler ridiculement fausses.

La nouvelle approche de Google : le Knowledge Vault

Knowledge Vault* fuses all these signals together



Pour résoudre ce problème, des chercheurs de Google, avec à leur tête Kevin Murphy, ont imaginé une nouvelle approche pour compléter la base de faits existante en utilisant des données extraites du web. Cette approche fait encore largement appel à Freebase, mais pour calculer une évaluation de la confiance que l'on peut accorder aux données extraites du web. C'est cette approche qu'ils ont appelé le Knowledge Vault, par opposition au Knowledge Graph, dans un article scientifique publié fin 2013.

Cette publication a été remarquée par la revue New Scientist, qui a publié un papier fort remarqué sur le sujet. Mais l'article du New Scientist laisse penser que les travaux sont terminés, les solutions trouvées et que le « Knowledge Vault » va bientôt supplanter le « Knowledge Graph ». La réalité est légèrement différente...

Kevin Murphy a communiqué quelques indications sur la volumétrie des données récupérées grâce à cette méthode fin 2013. Selon lui, il s'agit de la plus grande base de faits extraits automatiquement jamais collectée. Et cette base est dix fois plus grande que les expériences déjà menées à bien.

Mais en réalité, le volume des données extraites du Web étaient encore en deçà des résultats obtenus pour le Knowledge Graph. On peut certes imaginer que le travail s'est poursuivi et qu'aujourd'hui, ces seuils ont été dépassés, mais l'équipe de Kevin Murphy ne prétend pas avoir trouvé LA solution pour développer de manière sérieuse la base de faits utilisée par Google.

C'est une étape, et non un aboutissement

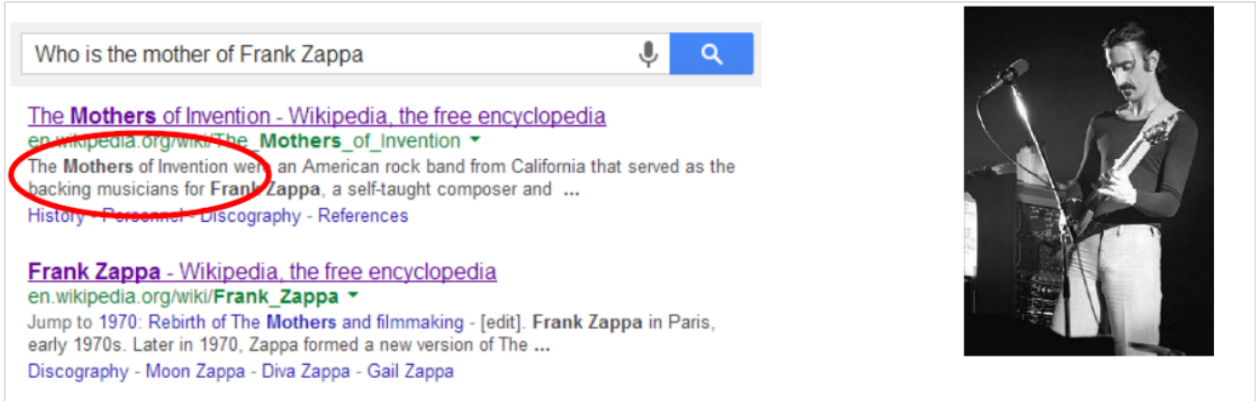
Au contraire, dans une conférence délivrée dans le cadre du séminaire CKIM fin 2013, Kevin Murphy a listé toutes les pistes qu'ils comptaient encore exploiter pour améliorer la couverture de ce système. Et sur ces points, ils en étaient encore à tester l'approche et la qualité comme la quantité de résultats obtenus.

Interroger le web

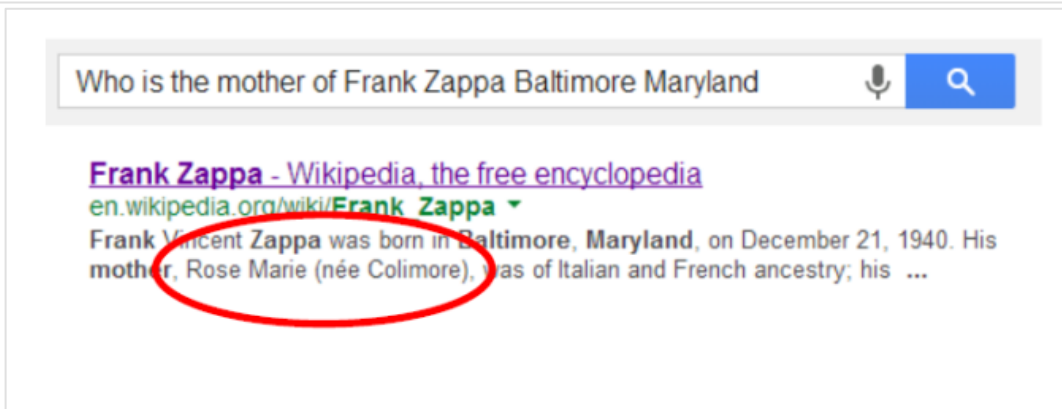
L'une des pistes pour extraire des faits non identifiés par des méthodes classiques, consiste à utiliser... le moteur de recherche pour les identifier. Une requête appropriée permet de faire

remonter une page qui contient l'information. Cette technique reproduit ce que les internautes faisaient avec un moteur navigationnel : chercher les URL qui sont susceptibles de contenir l'information.

La (très grande) difficulté ici est d'identifier la requête qui fera remonter l'information. Par exemple, la requête « Who is the mother of Franck Zappa » ne fait pas remonter de pages contenant la bonne information. Mais la requête « Who is the mother of Frank Zappa Baltimore Maryland » fait remonter l'information exacte, parce que Baltimore Maryland est son lieu de naissance.



The screenshot shows a search engine interface with the query "Who is the mother of Frank Zappa". The search results list two Wikipedia entries. The first entry, "The Mothers of Invention - Wikipedia, the free encyclopedia", has its title and the first sentence of the abstract circled in red. The second entry, "Frank Zappa - Wikipedia, the free encyclopedia", is also visible. To the right of the search results is a black and white photograph of Frank Zappa playing an electric guitar on stage.



The screenshot shows a search engine interface with the query "Who is the mother of Frank Zappa Baltimore Maryland". The search results list one Wikipedia entry, "Frank Zappa - Wikipedia, the free encyclopedia". The title and the first sentence of the abstract are circled in red. The first sentence reads: "Frank Vincent Zappa was born in Baltimore, Maryland, on December 21, 1940. His mother, Rose Marie (née Colimore), was of Italian and French ancestry; his ...".

Interroger les gens

L'autre piste, c'est d'étendre la logique de crowdsourcing en interrogeant les internautes pour découvrir les faits qui ne peuvent pas être extraits du Web (soit parce techniquement la donnée n'a pas pu être extraite, soit parce qu'elle n'était pas présente sur une page web). L'idée principale dans ce cas est de solliciter des feedbacks d'utilisateurs ou de les inciter à « compléter » Freebase. Une autre piste serait, *via* un lien sponsorisé, de diriger des internautes vers un « quiz ». Mais il faut aussi tenir compte de la « fiabilité » de ces fournisseurs d'information.

Cette méthode suppose donc que l'on trouve une méthode efficace pour évaluer les contributeurs, ce qui n'est pas tout à fait trivial.

Les principaux problèmes non encore totalement résolus

Pour couronner le tout, Kevin Murphy liste les défis qui restent à relever avant de disposer d'un système d'extraction universel et fiable :

- **Les entités manquantes** : la volumétrie des entités non détectées est encore considérable.
- **Les relations manquantes** : les méthodes actuelles sont surtout incapables d'extraire toutes les relations utiles.

- **Les informations implicites** : on trouve souvent des textes qui contiennent des faits qu'un humain saura extraire, mais pas une machine, parce que le « fait » n'est pas explicite, mais implicite : il découle du contexte. Par exemple, sur un site dédié à Victor Hugo, la page « bibliographie » contiendra potentiellement beaucoup de faits intéressants, mais si le terme « Victor Hugo » ne figure pas explicitement en regard de chaque livre (ce qui est probablement le cas) un extracteur automatique passera à côté de la relation « livres écrits par Victor Hugo ».

- **L'évaluation de la qualité des sources** : les méthodes d'extraction automatiques de faits retournent trop de données erronées. N'utiliser que des sources fiables serait évidemment un progrès mais les méthodes infaillibles pour identifier ces sources restent à inventer.

- **La prise en compte des contextes fictionnels** : l'un des principaux pièges que recèle l'extraction de données à partir du web est que dans certains cas, le fait n'est pas réel car la page parle d'une « fiction ». Or le web est rempli de sites qui reprennent ces informations fictionnelles. Par exemple, un système d'extraction automatique de faits pourra croire que la profession d'Abraham Lincoln, c'est d'être chasseur de vampires, parce que de nombreux sites parlent du film « Abraham Lincoln Vampire Hunter ».

Qu'est-ce que le Knowledge Vault peut changer ?

Néanmoins, à terme, les recherches en cours vont forcément aboutir à un changement d'échelle dans la volumétrie et la diversité des faits et relations collectées. Cela signifie d'une part qu'une proportion très importante de requêtes informationnelles pourraient recevoir une réponse directe (et diminuer encore le trafic envoyé depuis les moteurs de recherches aux sites web). Et d'autre part que l'on pourrait assister à une amélioration sensible du confort d'utilisation de la recherche vocale (de type Siri, Voice Search, Cortana, et qui suppose que l'on soit en mesure de renvoyer une info directe à l'internaute, et non une liste de liens bleus). Les bases de faits sont également très utiles pour améliorer sensiblement des applications de type « Google Now ».

De quelle échéance parle-t-on ici ? En fait, l'élargissement des bases de connaissance sera progressif, il ne faut pas s'attendre à des sauts quantiques, mais plus probablement à un processus graduel d'amélioration.

Du coup, qu'est-ce que cela implique en matière de SEO ? Comme ces applications de type « moteurs de réponse » court-circuiteront la visite de votre site s'il contient aujourd'hui des informations susceptibles d'être extraites, ce sont des typologies entières de sites qui reposent aujourd'hui sur le SEO qui risquent de recevoir beaucoup moins de trafic, au risque de remettre en cause des modèles économiques entiers (annuaires, bases encyclopédiques, bases de données en ligne, etc.). Faciliter l'extraction des informations sur votre site ne fera évidemment qu'accentuer le phénomène à votre détriment. Le gêner ne solutionnera pas forcément le problème si l'information que vous cherchez à garder pour vous et vos visiteurs est facilement extractible sur le site d'un de vos concurrents...

Mais il faudra au moins quelques années avant que les usages de recherches aient évolué de façon suffisamment notable pour que le fonctionnement en « moteur de réponses » ait sérieusement supplanté le « moteur navigationnel ». En attendant, l'impact négatif sur le trafic SEO de certains sites commence à être observable... Et on peut prédire que le phénomène ne fera qu'augmenter.

Bibliographie

L'article du New Scientist

Google's factchecking bots build vast knowledge bank

http://www.newscientist.com/article/mg22329832.700-googles-factchecking-bots-build-vast-knowledge-bank.html?full=true#.U_tDckjPbSC

La présentation de Kevin Murphy au CKIM 2013

From big data to big knowledge

<http://cikm2013.org/slides/kevin.pdf>

Publications scientifiques

Knowledge Vault: A Web-Scale Approach to Probabilistic Knowledge Fusion

Xin Luna Dong, Evgeniy Gabrilovich, Jeremy Heitz, Wilko Horn, Ni Lao, Kevin Murphy, Thomas Strohmann, Shaohua Sun, Wei Zhan, Google, 1600 Amphitheatre Parkway, Mountain View, CA 94043

<http://www.cs.cmu.edu/~nlao/publication/2014.kdd.pdf>

Open information extraction from the web.

Banko, M., Cafarella, M., Soderland, S., Broadhead, M., Etzioni, O.

In: Proceedings of the 20th International Joint Conference on Artificial intelligence, Hyderabad, pp. 2670–2676. Morgan Kaufmann, San Francisco (2007)

<http://turing.cs.washington.edu/papers/ijcai07.pdf>

The tradeoffs between open and traditional relation extraction.

Banko, M., Etzioni, O

In: Proceedings of ACL-08: HLT, Columbus, pp. 28–36. Association for Computational Linguistics, Columbus (2008)

<http://turing.cs.washington.edu/papers/acl08.pdf>

Regroupement sémantique de relations pour l'extraction d'information non supervisée

Wei Wang, Romaric Besançon, Olivier Ferret, Brigitte Grau

<http://www.taln2013.org/actes/www/TALN-2013/actes/taln-2013-long-026.pdf>

Open Information Extraction: the Second Generation

Oren Etzioni, Anthony Fader, Janara Christensen, Stephen Soderland, and Mausam

<http://turing.cs.washington.edu/papers/etzioni-ijcai2011.pdf>

Philippe YONNET, *Directeur Général de l'agence Search-Foresight, groupe My Media*
(<http://www.search-foresight.com>).

Référencement, vie privée et droit à l'oubli

[Retour au sommaire de la lettre](#)

Domaine :	Recherche	Référencement
Niveau :	Pour tous	Avancé

Cela fait plusieurs mois que l'arrêt de la Cour européenne de justice a obligé Google à mettre en place une procédure de droit à l'oubli pour certains documents, à la demande de personnes s'estimant lésées. Cet article fait le point sur les aspects juridiques de cette décision ainsi que sur ses conséquences, tout comme sur la façon dont Google prend finalement en compte ces demandes...

Le droit à l'oubli, propre aux personnes physiques, est inscrit en droit français depuis plus de 36 ans et en droit européen depuis près de 20 ans et pourtant, il a fallu plusieurs années pour que ce droit s'impose aux opérateurs d'Internet, à commencer par les moteurs de recherche. En effet, les autorités nationales ont appliqué les principes généraux de la protection des données personnelles à Internet dans les années 90 pour, doucement mais sûrement, affiner ces principes et détailler l'ampleur des obligations des moteurs. Le droit à l'oubli a été réaffirmé pour les moteurs de recherche très récemment par la Cour de Justice de l'Union européenne, notamment par une obligation de déréférencer tout contenu visant des données personnelles pour lesquelles les personnes concernées se sont manifestées. Retour sur cette nouvelle règle et son application...

Les principes généraux de la réglementation sur les données personnelles

Le droit européen prévoit que toute personne peut voir ses données personnelles traitées sous réserve que celles-ci soient (a) collectées pour des finalités déterminées, explicites et légitimes, et ne pas être traitées ultérieurement de manière incompatible avec ces finalités, (b) traitées de manière adéquate, pertinente et non excessives au regard des finalités pour lesquelles elles sont collectées et pour lesquelles elles sont traitées ultérieurement et (c) conservées sous une forme permettant l'identification des personnes concernées pendant une durée n'excédant pas celle nécessaire à la réalisation des finalités pour lesquelles elles sont collectées ou pour lesquelles elles sont traitées ultérieurement (extraits de l'article 6 de la directive 95/46/CE).

En d'autres termes, le responsable du traitement doit veiller à ce que le délai de conservation soit compatible et non excessif au regard des finalités pour lesquelles les données ont été collectées.

De plus, les personnes dont les données personnelles sont collectées ont des droits spécifiques, dont la possibilité de procéder à la rectification, l'effacement ou le verrouillage des données dont le traitement n'est pas conforme à la directive (en ce compris notamment le délai de conservation) (article 12) et de s'opposer, dans certains cas mais à tout moment, pour des raisons prépondérantes et légitimes tenant à sa situation particulière, à ce que des données la concernant fassent l'objet d'un traitement (article 14). Ces droits sont connus sous le terme du « droit d'opposition et de rectification » qui couvre également la conservation des données. En d'autres termes, si le « droit à l'oubli » n'a pas été expressément visé, il est induit dans la réglementation européenne.

L'arrêt du 13 mai 2014 de la CJUE et les nouvelles règles imposées aux moteurs

C'est sur le fondement de cette directive (qui reste, à ce jour, le seul texte européen relatif à la protection des données personnelles) que la Cour de Justice de l'Union européenne (cour suprême de l'Union européenne) a « découvert » le droit à l'oubli numérique. En effet, la Cour a un rôle certes limité dans l'interprétation et l'application de textes européens, mais comme toute cour suprême, elle peut parfois avoir un esprit créatif et aller plus loin que le texte lorsque les circonstances le permettent.

Nous avons évoqué le cas ayant conduit à l'arrêt du 13 mai 2014 dans la Lettre "Recherche & Référencement" de Juin 2014. Pour simple rappel, ce cas visait un entrepreneur qui avait été dans l'obligation de vendre aux enchères un bien immobilier suite à une saisie en 1998. Lorsqu'un internaute introduisait son nom dans Google, la liste de résultats affichait des liens vers deux pages du quotidien de la Vanguardia, datées de janvier et mars 1998 qui annonçaient notamment la vente destinée à recouvrer les dettes de sécurité sociale dues par M. González. Cet entrepreneur a demandé à la CNIL espagnole (AEPD) d'enjoindre le quotidien de désindexer cette information et d'enjoindre Google de faire de même soutenant que la saisie dont il avait fait l'objet avait été entièrement réglée depuis plusieurs années et que la mention de celle-ci était désormais dépourvue de toute pertinence. L'AEPD a demandé à Google de prendre les mesures nécessaires pour retirer les données de leur index. Google a immédiatement saisi la justice pour demander l'annulation de la décision de l'AEPD. C'est dans ce contexte que la juridiction espagnole a déféré une série de questions à la Cour de justice européenne.

La Cour a constaté qu'il y a lieu de rechercher un **juste équilibre** notamment entre un droit à l'information / accès à l'information propre à Internet et le respect de la vie privée et le droit à la protection de données à caractère personnel. La Cour a relevé que cet équilibre peut toutefois dépendre, dans des cas particuliers, de la nature de l'information en question et de sa sensibilité pour la vie privée de la personne concernée ainsi que de l'intérêt du public à recevoir cette information, lequel peut varier, notamment, en fonction du rôle joué par cette personne dans la vie publique. En d'autres termes, la Cour a laissé un certain nombre de cas dans lesquels un moteur peut refuser de répondre positivement à des demandes de personnes physiques.

La Cour précise que la personne concernée peut adresser de telles demandes directement à l'exploitant du moteur de recherche qui doit alors dûment examiner le bien-fondé de celles-ci. Lorsque le responsable du traitement ne donne pas suite à ces demandes, la personne concernée peut saisir l'autorité de contrôle ou l'autorité judiciaire pour que celles-ci effectuent les vérifications nécessaires et ordonnent à ce responsable des mesures précises en conséquence.

L'avancé de la jurisprudence française dans le sillage de la décision européenne

Cette décision européenne semble conforter la justice française dans ses décisions en donnant aux juges nationaux un cadre plus clair et un fondement juridique moins sibyllin. A ce titre, l'arrêt de la Cour de cassation du 10 septembre 2014 peut être souligné en ce que la cour suprême française a refusé de sanctionner le fait de simplement mettre en mot-clé le nom d'une personne sans y associer d'autres éléments répréhensibles.

Dans le cas de cette affaire, une personne - manifestement en procès pour diffamation avec des tiers - a mis sur son blog des informations personnelles concernant ses adversaires, en mettant dans les meta-tags, uniquement les noms de ces protagonistes. Les adversaires concernés ont saisi la justice, mais étrangement, avec de mauvais arguments et un fondement juridique sans rapport avec la loi « informatique et libertés » (alors que c'est cette loi qui régit les données personnelles). La Cour suprême a donc logiquement écarté cette argumentation étrange en expliquant que le choix d'un nom comme mot clé n'est pas constitutif d'une faute lorsqu'il n'y a pas d'autres éléments associés cherchant à nuire à la personne ou intrinsèquement répréhensibles.

Au-delà du fait que cet arrêt apporte peu sur la thématique du droit à l'oubli numérique parce que les arguments avancés étaient inappropriés, il faut souligner que depuis quelques mois,

les tribunaux et cours n'hésitent pas à mettre en œuvre ce droit, mais surtout à appliquer les limitations de ce droit. Limites que Google principalement a défini sur la base des termes de l'arrêt de la CJUE.

Comment Google met en œuvre la décision de la CJUE

Google a immédiatement réagi en mettant en ligne une page de demande de suppression de données personnelles (https://support.google.com/legal/contact/lr_eudpa?product=websearch&hl=fr). Il y est bien précisé que les demandes doivent émaner des personnes directement ou d'un représentant (comme un avocat), avec la copie d'une pièce d'identité ainsi que les URL incriminées.

Le 10 octobre 2014, Google a annoncé avoir reçu près de 145.000 demandes d'internautes européens pour effacer certaines de leurs informations, en un peu plus de quatre mois. Dès le premier jour, Google avait reçu 12.000 demandes en ce sens. Depuis, ce nombre a grimpé à 144.954 réclamations dont 28.912 demandes en France, qui ont provoqué la suppression de 497.695 pages web du moteur de recherche, précise le rapport. Google a cessé d'indexer 58,2% des liens qui lui ont été signalés.

Les moteurs déterminent la recevabilité d'une demande de suppression des moteurs. Concrètement, personne ne connaît avec précision les critères à ce jour et il est probable que les services internes juridiques ne les aient pas totalement définis. Il est donc possible de voir, malgré la mise en place de cette procédure, de nombreuses personnes déférer devant la CNIL (ou équivalent ailleurs en Europe), les décisions des moteurs de ne pas déréférencer certains contenus. Aujourd'hui, en toutes hypothèses, c'est aux moteurs et principalement à Google, de prendre de telles décisions, d'ordre juridique, alors que ces décisions devraient relever de la justice.

Sur la page dédiée à ses rapports de transparence portant sur les désindexations demandées par des tiers notamment au titre du droit à l'oubli numérique (<https://www.google.com/transparencereport/removals/europeprivacy/?hl=fr>), Google donne des exemples de désindexation. Par exemple, « Une personne nous a demandé de supprimer les liens vers des articles sur Internet qui font référence à son licenciement pour des crimes sexuels commis dans le cadre de son activité professionnelle. Nous n'avons pas supprimé les pages des résultats de recherche. »

L'étude des cas permet de délimiter plusieurs cas de raisons pour lesquelles Google peut être amenée à dire « Non ».

Répartition des « Non » par type de refus	Type de refus	% d'URL concernées
Concerne votre vie professionnelle		26%
Vous êtes à l'origine de ce contenu		22%
Toujours d'actualité et d'intérêt public		14%
Il s'agit de votre profil de réseau social		13%
L'information concerne une autre personne		7%
Pertinent, d'actualité et d'intérêt public		6%
Vous êtes un personnage public		4%
Ne concerne pas une personne physique		4%
Données gouvernementales		2%
Litige		1%
Autres		1%
Total		100%

De plus, Google souligne, sans surprise, que les sites les plus affectés sont des sites communautaires ou réseaux sociaux (même si on s'attendait à voir plus souvent Twitter).

Sites les plus affectés

La liste ci-dessous met en évidence les domaines pour lesquels nous avons supprimé le plus d'URL au sein de nos résultats de recherche.

www.facebook.com URL supprimées : 3353	profileengine.com URL supprimées : 3298	www.youtube.com URL supprimées : 2397	badoo.com URL supprimées : 2206	groups.google.com URL supprimées : 1949
www.yasni.de URL supprimées : 1559	www.wherevent.com URL supprimées : 1514	www.192.com URL supprimées : 1412	www.yasni.fr URL supprimées : 1299	www.yatedo.fr URL supprimées : 1174

Les délais de traitement sont très variables, mais sont globalement satisfaisants car inférieurs au mois en général.

Si une décision de Google n'est pas satisfaisante, la personne concernée peut saisir l'autorité locale (la CNIL en France) qui a compétence pour demander à Google de revoir sa décision.

L'expérience nous démontrera dans les années à venir si les moteurs savent garder raison ou au contraire abuser de ce pouvoir exorbitant (et incroyable dans une démocratie !) où la justice a accepté de se décharger sur un acteur privé. Il est probable que de nouveaux désaccords et contentieux naissent de cette nouvelle fonction attribuée aux moteurs. La partie n'est certainement pas terminée...

Alexandre Diehl

Avocat à la Cour, cabinet Lawint (<http://www.lawint.com/>)

Revue d'Outils Moteurs et SEO

[Retour au sommaire de la lettre](#)

Domaine :	Recherche	Référencement
Niveau :	Pour tous	Avancé

Cette rubrique signale chaque mois 5 outils (site web, logiciel, plug-ins Firefox...) qui nous ont parus intéressants (ou parfois étonnants...) pour le référencement naturel ou la recherche d'information. Des outils pas obligatoirement très récents, mais parfois méconnus ou qui gagnent à être utilisés selon nous. Chaque outil est décrit en quelques phrases, sous la forme d'une fiche signalétique vous permettant de le tester au plus vite...

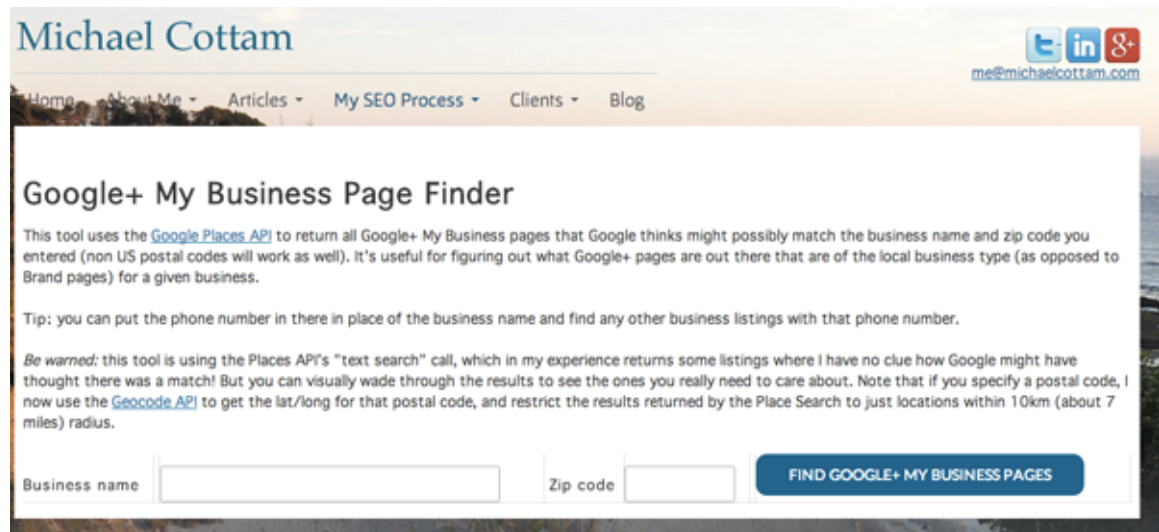
Nom : Moteur My Business

URL : <http://www.michaelcottam.com/google-business-page-finder/>

Domaine : Recherche

Type : Moteur de recherche de sites Google MyBusiness

Descriptif : Un moteur de recherche qui permet d'identifier des pages Google+ MyBusiness en saisissant un code postal, un nom d'entreprise, etc. Intéressant malgré quelques lacunes expliquées ici : <http://www.visibilite-referencement.fr/blog/un-moteur-de-recherche-de-page-my-business-google-205>



Nom : SEO Cleaner

URL : <http://www.seo.fr/seo-cleaner/>

Domaine : SEO

Type : Logiciel nettoyeur de backlinks

Descriptif : explication par le site lui-même : *SEO Cleaner a été créé par des référenceurs pour des référenceurs. C'est un crawler web conçu pour analyser les backlinks d'un site. Véritable Data Miner, SEO Cleaner va analyser les pages des liens qui lui sont soumis pour en*

Nom : SEO Site Checkup

URL : <http://seositecheckup.com/> et <http://suite.searchmetrics.com/en/research>

Domaine : SEO

Type : Sites web

Descriptif : 2 outils en ligne qui proposent un audit SEO de site web (contenu, backlinks, etc.) A tester et à comparer aux dizaines d'autres similaires existant déjà sur le Web.



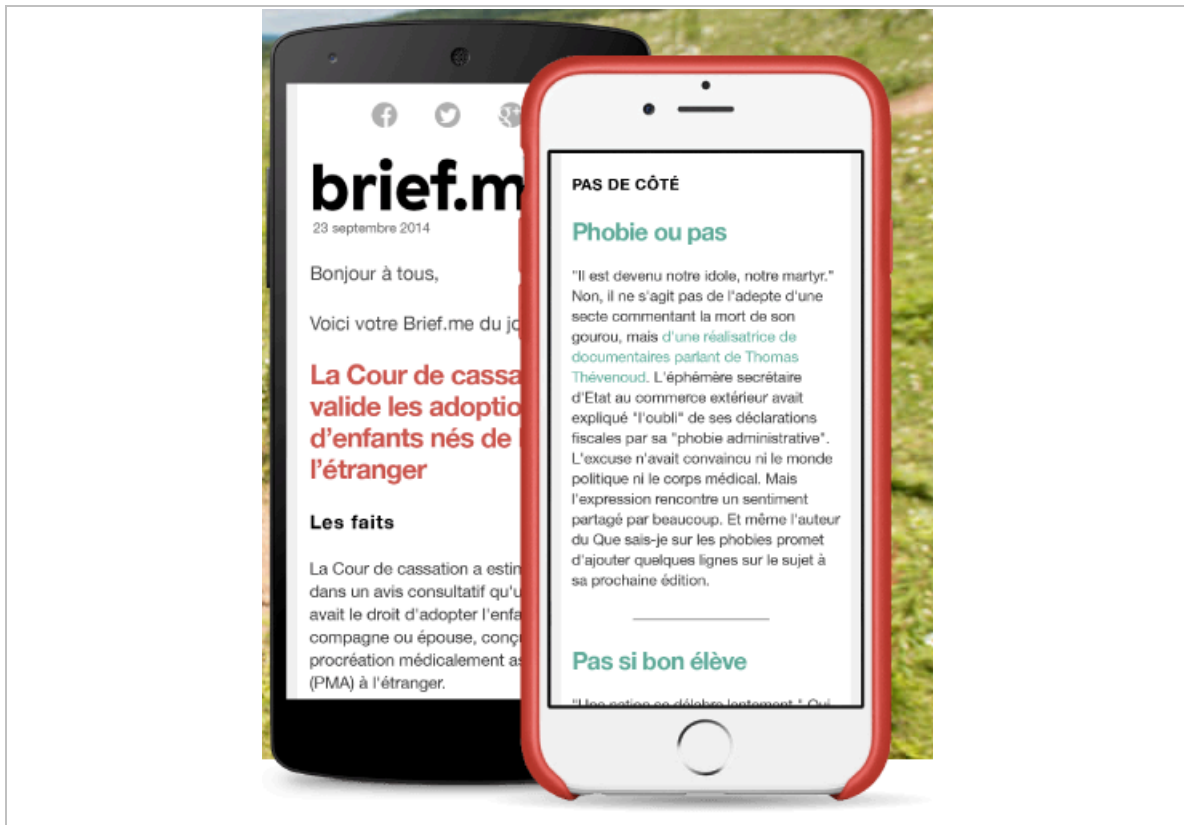
Nom : Brief.me

URL : <http://www.brief.me/>

Domaine : Recherche

Type : Moteur de recherche d'actualités

Descriptif : "Brief.me explique ce qui est compliqué, résume ce qui est long, analyse ce qui est important." Le site (en construction) est actuellement en phase de crowdfunding. A suivre de près...



Rubrique proposée par Olivier Andrieu, éditeur du site Abondance.com.

Revue d'URL

[Retour au sommaire de la lettre](#)

Domaine :	Recherche	Référencement
Niveau :	Pour tous	Avancé

Les articles de notre veille sur les moteurs de recherche et le référencement en français et en anglais, ainsi que le dossier du mois (lorsque l'actualité fait ressortir un fait marquant)...

Revue d'URL en français...

-> [Flash Référencement 22] Comment réagir à une mise à jour d'algorithme Google ?
Quand Google lance une mise à jour d'algo' : « Panic or not » ? C'est l'algorithme de google qui détermine le classement des chaque site pour une requête donnée, il est donc naturel de s'inquiéter (ou de se réjouir) lorsque les sites d'actu SEO annoncent une mise à jour en cours...

<http://www.yakamama.com/comment-reagir-mise-a-jour-algorithme-google>

-> Les postes les plus rémunérateurs chez Google
Google adore ses techniciens. Il n'est donc pas très surprenant que les ingénieurs logiciels fassent partie des employés les mieux payés de la société...

<http://www.journaldunet.com/ebusiness/le-net/salaires-google/>

-> Infographie : 16 conseils pour créer du contenu intelligent pour votre site web
Dans le milieu du web, tout le monde est aujourd'hui convaincu que le contenu est roi et que c'est par du contenu de haute qualité qu'il faudra passer pour ranker dans les premières positions du moteur de recherche. Mais créer du contenu à haute valeur ajoutée n'est pas donné à tout le monde. Voici donc quelques conseils qui vous permettront de proposer du contenu attrayant...

<http://www.referenceur.be/infographie-16-conseils-creer-contenu-site-web-13599.html>

-> Les mots clés de WordPress et référencement naturel
En référencement naturel, il faut savoir optimiser son site, mais encore faut-il savoir correctement l'utiliser. Et sur WordPress, certains éléments peuvent être traîtres, comme par exemple les fameux « Mots clés »...

<http://www.seomix.fr/problemes-mots-cles-wordpress-en-referencement/>

-> Netlinking qualitatif

Le VLC 2014 commence à poindre le bout de son nez. Raph nous jette quelques miettes d'infos par ci par là histoire de nous mettre l'eau à la bouche. Histoire de clore définitivement l'aventure de l'an dernier voici les slides de ma conf pour patienter en attendant le VLC NOUVEAU !!...

<http://gain-de-temps.com/blog/netlinking-qualitatif/>

-> Ressources SEO & Online Marketing (Google proof)

Je vous propose de trouver ici une sélection de ressources SEO & autres tirées de mes favoris, toutes liées à la visibilité des sites web et que je trouve particulièrement utiles et pertinentes (il y a beaucoup d'articles en Anglais par contre -on ne choisit pas, que voulez-vous)...

<http://seofactorygirl.com/ressources/>

-> Interview Olivier Andrieu, le meilleur référenceur Français

Aujourd'hui je vous présente cette belle interview de Olivier Andrieu, le meilleur référenceur français, auteur du livre "SEO ZERO EURO". Vous allez tout savoir sur l'auteur, sur son livre et/ou presque tout, dans cette interview...

<http://test-seo-mca.blogspot.fr/2014/09/interview-olivier-andrieu-le-meilleur.html>

-> Marissa Mayer, depuis la geek de Google à la CEO de Yahoo

Dans le monde masculin du web et des informaticiens, Marissa Mayer, CEO de Yahoo, fait figure d'exception. Elle a su se faire une place et imposer son style...

<http://www.journaldunet.com/ebusiness/le-net/marissa-mayer-marissa-mayer-biographie.shtml>

-> Le web social ce n'est pas magique

J'ai une forte propension à croire que tout le fatras d'outils disponibles pour communiquer sur le web est une véritable chance pour les entreprises d'améliorer leur visibilité et leur relation avec leurs prospects et leurs clients...

<http://blog.axe-net.fr/le-web-social-ce-nest-pas-magique/>

-> Voici les avantages préférés des salariés de Google

Les postes chez Google font partie des plus convoités de toute l'industrie technologique. En plus de bureaux absolument magnifiques, Google propose de nombreux avantages à ses employés. D'anciens Googlers et d'autres qui travaillent toujours dans la société ont listé leurs favoris sur Quora...

<http://www.journaldunet.com/ebusiness/le-net/avantages-des-salaries-chezgoogle/>

-> Evolution de la recherche sur le World Wide Web

Petit retour cette semaine sur une histoire de l'internet et des moteurs de recherche. La connaissance du passé nous fait mieux comprendre ce qui se déroule aujourd'hui et participe à une meilleure anticipation de ce qu'il va advenir ! Voilà pour la bonne parole :) ...

<http://www.mauricelarger.com/historique-de-la-toile-depuis-les-annees-90/>

-> Béatrice Payonne : Paroles de SEO

Café Référencement poursuit sa série d'interviews de référenceurs inhouse. Aujourd'hui, Béatrice Payonne du Crédit Agricole...

<http://www.cafe-referencement.com/actu-seo/beatrice-payonne-paroles-seo-1264>

-> 15 déclarations insolites d'Eric Schmidt, président de Google

Le président de Google est réputé pour n'avoir pas froid aux yeux. Il n'hésite en effet pas à être souvent téméraire dans ses déclarations. Jugez plutôt...

<http://www.journaldunet.com/ebusiness/le-net/eric-schmidt.shtml>

-> Cyril Arnaud sur le web-gril

Nous poursuivons notre série d'interviews sur le quotidien des pro du webmarketing français. Cette semaine, c'est Cyril Arnaud, détenteur du compte @actualites_seo qui se prête à l'exercice du web-gril ! Ce SEO lyonnais freelance défend dans ces quelques lignes ses méthodes de white hat convaincu...

<http://www.editorial-interactif.com/cyril-arnaud-sur-le-web-gril/>

-> Pourquoi ajouter du contenu à son site pour améliorer le référencement naturel d'un site ?

Vous avez un site web et vous voulez le rendre visible. Pour cela, on vous a dit qu'il fallait l'optimiser, lui ajouter du contenu et lui créer des liens. Nous allons nous concentrer sur la phase d'ajout de contenus en se préoccupant de 2 problématiques...

<http://www.agence-csv.com/blog/ajouter-du-contenu-site-ameliorer-referencement-naturel-dun-site/>

-> Anecdote e-commerce

Un peu de détente aujourd'hui, avec une mini histoire vécue dans le e-commerce...

<http://blog.axe-net.fr/anecdote-e-commerce/>

-> Google PageSpeed, vitesse de chargement des pages et nouveautés HTML 5

Le débat sur la vitesse de chargements des pages web existe depuis des années et l'arrivée du PageSpeed de Google. Les référenceurs ont toujours été partagés à ce sujet et estiment parfois que ce facteur n'est pas réellement pris en compte dans le positionnement des pages. Dans cet article, nous allons penser avant tout à nos utilisateurs et garder en trame de fond l'idée d'un meilleur référencement en voyant quelques innovations futures en HTML 5...

<http://blog.internet-formation.fr/2014/09/google-pagespeed-vitesse-de-chargement-pages-web-html5/>

-> Le référencement c'est mort ! Pas d'bol !

Vous n'êtes pas d'accord ? Comment ça ? Tout le monde le dit ! Alors oui, bien sûr comme c'est moi qui le dit, parce que je ne suis pas un cadavre du SEO, j'aurais tort ?!... Pfff !!! M'en fous, je sais que j'ai raison. Le référencement c'est mort, la loose je vous dis ! Aujourd'hui le nec plus ultra (de la mort qui tue), c'est le SMO. Et non, ça ne veut pas dire Sexual Mental Obsession (quoique, en cherchant un peu, on pourrait trouver une corrélation)...

<http://www.simplewebsite.fr/blog/referencement/le-referencement-cest-mort-pas-dbol/>

-> Interview – Danny Sullivan, LE gourou du search : « Google est archi-dominant, et pour longtemps »

Danny Sullivan est le grand gourou du search. Basé en Californie, il édite Searchengineland, la bible du search marketing, et organise les conférences SMX aux quatre coins de la planète...

<http://www.viuz.com/2014/09/30/interview-de-danny-sullivan-le-gourou-du-search-google-est-archi-dominant-et-pour-longtemps/>

-> Jean-Hugues Brretin, Yahoo! : Paroles de SEO

Café Référencement poursuit sa série d'interviews de référenceurs inhouse. Aujourd'hui, Jean-Hugues Bretin, SEO Manager France chez Yahoo!...

<http://www.cafe-referencement.com/actu-seo/jean-hugues-bretin-yahoo-paroles-seo-1314>

-> Qualité sémantique objective, les structures en SEO, opus 4

Vous avez eu plusieurs mois pour vous distraire, maintenant, il faut se remettre au boulot... On reprend les opus. Celui-ci, l'opus 4, va peut-être vous faire trembler (ce sera moins dur que le supplice de Tantale, rassurez-vous !)...

<http://www.referencement-naturel-white-hat.fr/qualite-semantique-objective-les-structures-en-seo-opus-4/>

-> Quelle stratégie de contenu adopter pour son site ?

Rédiger, rédiger, rédiger ... ces mots, vous les avez entendus maintes et maintes fois. On vous les rabâche dans tous les articles et à toutes les sauces : « votre site a besoin de contenus », « un site sans contenus ne sera jamais visible », « sans contenu, votre blog n'est rien ! »...

<http://www.pole-position-seo.com/2014/10/strategie-de-contenu-pour-son-site.html>

-> David Besbris : avec Google+, nous sommes encore là pour un bon moment !

Lors d'une interview sur recode.net, David Besbris le nouveau responsable de Google+ suite au départ de Vic Gundotra, a déclaré (ou confirmé) que Google+ n'est pas prêt de mourir et que le réseau social est encore là pour un bon moment...

<http://www.visibilite-referencement.fr/blog/david-besbris-google-nous-sommes-encore-la-pour-un-long-moment-203>

-> Le mythe des 200 facteurs de ranking de Google

Ohhhhh, mais en voilà un bien beau titre ! Bon ce n'est pas de moi, mais j'ai trouvé l'article original de Gianluca Fiorelli sur Moz fort sympathique et vous propose ici une adaptation francisée pour ceux qui sont fâchés avec la langue de Shakespeare. N'hésitez pas à me corriger si nécessaire...

<http://www.numastickwebfactory.fr/star-ducks-seo/mythe-200-facteurs-ranking-google.html>

-> Travailler chez Google : voici les critères qui comptent vraiment

La société est réputée pour poser toutes sortes de questions farfelues aux candidats qu'elle auditionne ; mais voici ce qu'elle recherche réellement...

<http://www.journaldunet.com/ebusiness/le-net/travailler-pour-google-les-criteres-qui-comptent.shtml>

-> Refonte d'une boutique en ligne, les points SEO à ne pas oublier

La refonte d'un site e-commerce peut se révéler une catastrophe sur le plan du trafic. Je travaille donc assez souvent pour des clients comme intermédiaire vis-à-vis des agences chargées de la mise en œuvre (sauf quand nous développons le e-commerce en interne bien sûr...)...

<http://blog.axe-net.fr/refonte-ecommerce-seo/>

Revue d'URL en anglais...

-> Google Rel=Agent Replaces Authorship: Maybe Google Testing SEOs
Some savvy webmasters were inspecting the HTML markup behind the last Google Webmaster Blog post and noticed the markup used on the posted by John Mueller hyperlink had a bit of extra code for rel=agent. Here is a screen shot of the code...

<https://www.seroundtable.com/google-rel-agent-19145.html>

-> Learning From Lost Link Opportunities
Instead of getting upset about rejection, take a few steps back and look at why you aren't getting the links you want...

<http://searchengineland.com/learn-lost-link-opportunities-202662>

-> Google Granted Patent On Using What You Watch On TV As A Ranking Signal
I'm going to have to turn up the sound on my TV, and decide carefully what to watch, and test this. It would be very interesting if it works. Is Google clued in to what you are watching on TV? If so, is that through a set top box, or an internet enabled television?...

<http://www.seobythesea.com/2014/09/google-patent-watch-tv-as-ranking-signal/>

-> Google's First Semantic Search Invention Was Patented In 1999
This is officially part of the story I'm telling in a presentation I prepared for SMX East, in a couple of weeks in New York. The name of the session I'm in is "Hummingbird and the Entity Revolution," which reminds me of a Prince song from the 1980s...

<http://www.seobythesea.com/2014/09/google-first-semantic-search-invention-patented-1999/>

-> How Google is Connecting Keyword Relevance to Websites through More than Just Domain Names - Whiteboard Friday

We're seeing Google continue to move beyond just reading pages, instead attempting to truly understand what they're about. The engine is drawing connections between concepts and brand names, and it's affecting SERPs. In today's Whiteboard Friday, Rand explains just what Google is doing, and how we can help create such associations with our own brands...

<http://moz.com/blog/google-keyword-relevance-whiteboard-friday>

-> How to use Keyword Research to Find New Landing Page Testing Ideas

Conversion rate optimization is all about finding the right elements on a page to test. Should you test a new value proposition, image, headline or redesign a page completely?...

<http://searchenginewatch.com/article/2371264/How-to-use-Keyword-Research-to-Find-New-Landing-Page-Testing-Ideas>

-> Links, FUD, and Webmaster Guidelines

The confusion and misinformation surrounding link-building is truly breathtaking. It's really no surprise, thanks to Google and the FUD they've spread. I was recently reminded about just how deep the misunderstanding runs, when I visited Google's Webmaster Central Help Forum dedicated to crawling, indexing, and ranking...

<http://searchenginewatch.com/article/2371510/Links-FUD-and-Webmaster-Guidelines>

-> How Google May Add To Its Knowledge Base With Entities And Attributes From Search Queries

Millions of searches stream into Google everyday as people try to meet their informational and situational needs. But those searches don't disappear after the searches. They provide Google with some very interesting and useful information in return. For instance, they tell Google what people are interested in real time - right at this moment...

<http://www.seobythesea.com/2014/09/google-may-add-knowledge-base-entities-attributes-search-queries/>

-> Recover From Panda? Follow These 5 Steps to Avoid Future Panda Hits

For those dealing with a major drop in traffic due to an algorithm update, I'm not sure there's a better feeling in the world than seeing recovery. The original algo hit typically feels like a punch in the gut, with serious confusion and frustration quickly following. Then you have months of recovery work with hope (and faith) that your hard work will pay off. And for some,

recovery will never happen. That's an unfortunate truth for sites that have pushed the limits of spamming repeatedly for a long time...

<http://searchenginewatch.com/article/2371778/Recover-From-Panda-Follow-These-5-Steps-to-Avoid-Future-Panda-Hits>

-> Google Turns To Deep Learning Classification To Fight Web Spam

In the past few years, Google has been busy building what has become known as the Google Brain team, which started out by having its deep learning approach watching videos until it learned to recognize cats...

<http://www.seobythesea.com/2014/09/google-turns-deep-learning-classification-fight-web-spam/>

-> Extracting Semantic Classes And Corresponding Instances From Web Pages And Query Logs

In creating a knowledge base, there seem to be a number of approaches that can be used to supply entities and facts from sources like web pages and query logs...

<http://www.seobythesea.com/2014/09/extracting-semantic-classes-instances-from-web-pages-query-logs/>

-> Using Query User Data To Classify Queries

It can be difficult classifying a query for a search engine based upon the query itself. For example, you could classify the query "lincoln" based upon...

<http://www.seobythesea.com/2014/09/queries-search-entities-query-user-data-classifications/>

-> Calculating The Risk Of Search Engine Spam

When was the last time you seriously considered what would happen if you lost all organic traffic to your site?...

<http://searchengineland.com/calculating-risk-search-engine-spam-204223>

-> New Panda Update; New Panda Patent Application

Google's Pierre Far announced on his Google+ page that Google was releasing a new Panda update that supposedly included some new signals that could potentially help "identify low-quality content more precisely."...

<http://www.seobythesea.com/2014/09/new-panda-update-new-panda-patent/>

-> The Myth of Google's 200 Ranking Factors

The woman in the gif below just said to Captain Picard that she can show him the definitive and complete list of the 200 Google ranking factors...

<http://moz.com/blog/the-myth-of-googles-200-ranking-factors>

-> Google: There Isn't A Magical SEO Advantage By Switching To HTTPS

Google's John Mueller said in a video hangout 22 minutes and 21 seconds in that there is no magical SEO advantage when switching from HTTP to HTTPS...

<https://www.seroundtable.com/google-https-rankings-19228.html>

-> How To Create A 301 Redirect Map

Overhauling your site? Without a 301 redirect map, you risk losing valuable search equity...

<http://searchengineland.com/create-301-redirect-map-203248>

-> At SMX East; Presenting On Google And The Semantic Web

The Semantic Web is making an even stronger appearance recently at Google than it has in the past. With knowledge panels, carousels listing all kinds of things (and people and places), structured snippets merging query answers with question answers into a single snippet, OneBoxes of many different kinds, and even Hummingbird responding better to longer and more complex queries, it's the future of Google...

<http://www.seobythesea.com/2014/10/google-semantic-web/>

-> Google On How They Know When To Slow Or Stop Crawling Your Web Site

Google crawl efficiency signals include connect time and HTTP server status codes...

<http://searchengineland.com/google-know-slow-stop-crawling-web-site-204674>

-> Google Acquires 55 Yahoo! Patent Filings

It was a surprise to see a number of Yahoo! patents listed in Google's assignment database as having been assigned to Google. With news recently that Yahoo would be closing the Yahoo Directory, that seemed like a strategic choice. Now I'm wondering if we will ever see an independent Yahoo Search Engine ever again once their deal to have Microsoft supply search results to them ends...

<http://www.seobythesea.com/2014/10/google-acquires-55-yahoo-patent-filings/>

-> 7-Point Relaunch SEO Checklist

Search engine optimization adherence in the website relaunch/redesign process is a topic I enjoy talking about. In the past few years I have covered a few topics of what to think about from an SEO perspective when undertaking a redesign effort and how to assess potential SEO issues you are experiencing after your new site has launched...

<http://searchenginewatch.com/article/2373910/7-Point-Relaunch-SEO-Checklist>

-> Panda 4.1: The Devil Is in the Aggregate

I wish I didn't have to say this. I wish I could look in the eyes of every victim of the last Panda 4.1 update and tell them it was something new, something unforeseeable, something out of their control. I wish I could tell them that Google pulled a fast one that no one saw coming. But I can't...

<http://moz.com/blog/panda-4-1>

-> What Do Structured Snippets Mean To Your Page-Level Content Strategy?

Google is now showing structured snippets in search results; Columnist Jim Yu shares tips to take advantage of this new format...

<http://searchengineland.com/structured-snippets-mean-page-level-content-strategy-204557>

-> 37 SEO Experts You Should Be Following on Twitter

I've been speaking at search conferences since 2001 and have spent a ton of time getting to know who is who in the industry. In fact, I look forward to seeing some of you at SES Chicago to discuss Buyer Legends. Every day it seems that there are another dozen experts to follow on SEO. But, who should you really be paying attention to in 2014?...

<http://searchenginewatch.com/article/2374336/37-SEO-Experts-You-Should-Be-Following-on-Twitter>

-> The Future of Link Building

Building the types of links that help grow your online business and organic search traffic is getting harder. It used to be fairly straightforward, back before Google worked out how to treat links with different levels of quality and trust. However, the fact that it's getting harder doesn't mean that it's dead...

<http://moz.com/blog/future-of-link-building>

-> Google May Add Mobile User Experience To Its Ranking Algorithm

Google sees what users see, and if that is a bad mobile experience, it may impact your rankings...

<http://searchengineland.com/google-may-add-mobile-user-experience-ranking-algorithm-205382>

-> Penguin Recovery: Should You Be Removing Links or Just Disavowing?

If you have been affected by Google's Penguin algorithm, should you be actively trying to remove your unnatural links? Or, is it just good enough to use Google's disavow tool? In this article we will discuss the pros and cons of each and also look at what Google has said on the subject...

<http://searchenginewatch.com/article/2374406/Penguin-Recovery-Should-You-Be-Removing-Links-or-Just-Disavowing>

-> Google SEO HTTPS Migration Checklist

Despite most not seeing many changes after migrating from HTTP to HTTPS for the Google HTTPS ranking boost, still webmasters want to do it...

<https://www.seroundtable.com/google-seo-http-to-https-migration-checklist-19268.html>

-> How to Acquire Anchor Text-Rich Links Without Resorting to Spam or Manipulation - Whiteboard Friday

All signs point to links with exact match anchor text retaining the huge value we've seen throughout the years, but many of the techniques for acquiring those links are spammy. There are a few, though, that not even Google would frown upon, and in today's Whiteboard Friday, Rand shows you what they are...

<http://moz.com/blog/anchor-text-rich-links-without-spam-or-manipulation-whiteboard-friday>

-> How Google May Answer Fact Questions Using Entity References In Unstructured Data

A Google patent application explores how Google may answer factual questions from unstructured Web pages and results rather than from more structured sources such as Freebase or Wikipedia. The processes described in the patent are pretty interesting, and they might be more familiar to an SEO trained audience than a Semantic Web one, like a result that ranks well because of a "query deserves freshness" approach...

<http://www.seobythesea.com/2014/10/google-fact-questions-entity-references-unstructured-data/>

-> Announcing the 2014 Local Search Ranking Factors Results

Many of you have been tweeting, emailing, asking in conference Q&As, or just generally awaiting this year's Local Search Ranking Factors survey results. Here they are!...

<http://moz.com/blog/local-search-ranking-factors-2014>

-> Penguin – What Happens Next? 10 Data-Led Predictions

Twitter has been alight for the past couple of weeks with news that the latest Penguin algorithm update is almost ready for action...

<http://searchenginewatch.com/article/2375404/Penguin-What-Happens-Next-10-Data-Led-Predictions>

-> Rich Snippets And Patterned Queries

Revisiting the Subscribed Links Patent Five Years Later and Finding the Rich Snippets Patent...

<http://www.seobythesea.com/2014/10/rich-snippets-patterned-queries/>

-> Dealing With Onsite Duplicate Content Issues

The problem of duplicate content arises when there is more than one version of a page indexed by the search engine. Duplicity can be both onsite and offsite: Onsite duplicity is when the same content is seen on multiple pages within a website and offsite duplicity is when the content on your website is similar to that on some other site...

<http://searchenginewatch.com/article/2375911/Dealing-With-Onsite-Duplicate-Content-Issues>

-> Google How To Snippets And Cooking With Semantics

Google recently started showing "How to" lists in search results, which tend to show the first few steps of some task, and then let you click through to a page to see more. Like the recipes above for things like guacamole...

<http://www.seobythesea.com/2014/10/google-snippets-cooking-semantics/>

Rubrique proposée par Olivier Andrieu, éditeur du site Abondance.com.

Merci pour votre lecture...

Pour toute suggestion au sujet de cette lettre professionnelle : olivier@abondance.com