

YourTextGuru, un nouvel outil d'aide à la rédaction



Par Guillaume et
Sylvain Peyronnet

Domaine :	Recherche	Référencement
Niveau :	Pour tous	Avancé

Pour rédiger des textes qui soient pertinents pour l'internaute et optimisés pour les moteurs de recherche, il est nécessaire d'avoir des repères. Une voie pour cela est l'utilisation de corpus, mais on se rend assez vite compte que cela peut mener à suroptimiser les rédactionnels. Une autre façon de faire est d'utiliser des guides de rédaction fournis automatiquement par des outils d'analyse du Web et des résultats de Google sur une requête ou un concept donné. L'outil YourTextGuru a cette ambition. Mis en œuvre par Guillaume et Sylvain Peyronnet, il se veut être, pour un prix modique, le compagnon au quotidien du rédacteur web. Explications sur son fonctionnement et la meilleure façon de l'utiliser.

En référencement web, on ne le répètera jamais assez, le contenu a toute son importance. Il devient de plus en plus difficile de positionner des pages web sur des requêtes qui n'ont rien à voir avec les contenus qui y sont évoqués. Les différentes surcouches plus récentes chez Google destinées à filtrer les contenus de mauvaise qualité n'y sont d'ailleurs pas totalement étrangères...

Dès lors, la qualité des contenus, évoquée depuis des années comme primordiale via la maxime « Content is King », devient d'autant plus essentielle. Et la question qui brûle toutes les lèvres, surtout celle des référenceurs, apparaît en filigrane : "Qu'est-ce qu'un contenu de qualité ?"

Cet article a pour ambition de donner un élément de réponse satisfaisant à cette question, en abordant une certaine manière de produire de tels contenus, en faisant appel à un nouvel outil appelé YourTextGuru.

Pour la petite histoire...

Il y a quelques mois, en tant qu'abonné-e-s à la Newsletter R&R, vous avez peut-être eu l'opportunité de lire un article de notre part sur les corpus (<http://recherche-referencement.abondance.com/2015/05/le-corpus-comme-guide-doptimisation-des.html>). Sobrement intitulé « Le corpus comme guide d'optimisation des contenus », celui-ci expliquait comment utiliser les statistiques pour obtenir une liste de mots à utiliser pour optimiser des contenus. On y parlait notamment de TF.IDF et du cosinus de Salton, afin d'expliquer comment le moteur de recherche percevait la pertinence d'un texte.

Suite à cet article, qui complétait à merveille une conférence « live » que nous avons eu l'occasion de donner un peu plus tôt, les retours furent nombreux et enthousiastes. Les plus bricoleurs commencèrent à mettre en place en interne des outils de génération de corpus, tandis que certains disposant de ressources moins techniques nous passèrent un coup de fil pour acheter des corpus.

Mais après quelques ventes, il fallut se rendre à l'évidence : utiliser un corpus n'était pas quelque chose de très aisé pour celui qui n'était pas averti...

mot	TF (min)	TF (max)	IDF	tf.idf (sum)
cuisine	11,8 %	100,0 %	1,21	75,41
cours	16,7 %	100,0 %	1,27	69,49
chef	5,6 %	100,0 %	1,93	30,04
paris	11,1 %	100,0 %	2,86	24,78
atelier	5,6 %	100,0 %	2,23	23,48
recettes	7,7 %	100,0 %	2,31	15,37
restaurant	8,3 %	100,0 %	3,08	14,03
gourmand	5,9 %	100,0 %	3,37	12,73

Fig.1. Exemple de corpus (8 mots classés par somme des tf.idf décroissante)

En effet, le rédacteur final (qui est rarement celui à l'origine de la demande de corpus) avait tendance, dans un premier temps, à regarder le corpus d'un drôle d'air, puis à se gratter la tête, pour finalement écrire son texte en prenant soin d'y faire figurer le plus possible de mots en haut de la liste... Bien sûr, cela donnait des textes meilleurs que ceux écrits en improvisant. Mais cela donnait surtout des textes suroptimisés du point de vue référencement !

En effet, le corpus permet de connaître « l'importance » d'un mot pour une requête ou une thématique, mais cela ne veut pas dire, et c'est même l'inverse, qu'il faut absolument placer tous les mots « importants » de la liste. C'est contre-productif : au lieu de flirter avec la ligne jaune, on la dépasse, on la survole, et on passe même parfois directement dans la stratosphère !

Pensez au filtre Penguin qui pénalise les sites ayant trop de liens à ancrs exactes pointant vers lui : quelques ancrs exactes, c'est parfait, c'est même une belle optimisation SEO, ça marche très bien pour améliorer le positionnement sur la requête correspondant à l'ancre. Mais dès lors qu'on pousse l'optimisation un peu trop loin, on franchit la barrière. On passe de bon élève à cancre, voire tricheur, au yeux de Google.

Il en est de même avec les contenus. Bien sûr, il existe des sites de référence qui sont encyclopédiques et contiennent tout le dictionnaire de A à Z. Mais la plupart fournissent plutôt des textes répondant à des questions précises et ne sont pas destinés à répondre à la question de la vie, l'univers et le reste.

Quand on analyse les pages que l'on trouve dans les SERP en bonnes positions pour de nombreuses requêtes, on se rend compte que c'est exactement cela. La plupart des pages ne contiennent pas tout le savoir du monde. Elles sont optimisées, quand il y a une problématique SEO liée, mais rarement suroptimisées en termes de contenus. On y trouve un savant mélange entre mots communs, usuels et mots rares, experts et entités nommées.

C'est là tout l'enjeu de la rédaction pour le référencement web. Il faut parvenir à équilibrer le vocabulaire utilisé pour rester sous le radar du censeur.

Suite aux déconvenues des premiers utilisateurs de corpus, le choix fut rapidement fait de tout simplement arrêter de proposer d'en réaliser pour des clients. Le format corpus n'est pas un format très adapté pour le référenceur, encore moins pour le responsable du marketing ou encore le rédacteur final. En utilisant un corpus, difficile d'avoir un bon ROI, comparé au prix de vente commun d'un corpus (plusieurs centaines d'euros à l'époque).

YourTextGuru : Kezaco ?



Fig.2. Logo de l'outil YourTextGuru

Après quelques mois de réflexion et un échange, suite au « live » sur les corpus évoqué plus haut, avec Frédéric Bobet (TriKaya) et Benoit Chevillot (teknSEO), nous avons décidé, tous ensemble, de proposer une solution au problème des corpus. C'est-à-dire de garder la philosophie de base :

- Aider le rédacteur (web) à rédiger ;
- De façon statistiquement naturelle ;
- Tout en lui économisant du temps ;
- En n'oubliant pas que l'expertise est avant tout chez le rédacteur (web).

Afin d'améliorer le ROI pour le client, il fallait aussi penser à réduire drastiquement le coût de création d'un tel guide. Seule l'automatisation permettrait ce tour de force (et donc c'est un service web qu'il fallait mettre en place, pas un système de prestations). Et il faut bien dire qu'au moment du lancement de l'outil, avoir pu afficher un prix au guide de quelques dizaines de centimes d'euros est vraiment quelque chose dont nous sommes fiers.

Algorithmes et technique

Sans aller jusqu'à dévoiler toutes les petites ficelles derrière YourTextGuru (accessible à l'adresse <https://yourtext.guru>), on peut révéler beaucoup de choses. D'abord, du point de vue algorithmique, finie la notion de corpus, de TF.IDF, de cosinus de Salton, etc. Un

guide de rédaction doit permettre d'écrire un texte. C'est donc un modèle génératif de la langue que l'on a souhaité mettre en oeuvre plutôt qu'une simple détection des éléments « importants ». C'est-à-dire, pour dire les choses simplement, que l'on observe des données et qu'on en tire des statistiques d'utilisations, à partir desquels on génère aléatoirement un « gabarit » de texte. Ce gabarit est unique, statistiquement naturel et très utile pour rédiger vite. Nous verrons un exemple de gabarit un peu plus tard dans cet article.

Au niveau technique, il existe plusieurs briques, écrites dans des langages variés (Python, Ruby, PHP, NodeJS, R). Celles-ci répondent parfois à un besoin très fonctionnel (lire des contenus sur le web, récupérer les résultats d'un moteur de recherche, servir les pages web du service, etc.), parfois à un besoin plus calculatoire (génération des statistiques, calcul de cooccurrences, génération des guides, etc.). Et bien sûr, qui dit data, dit entrepôts de stockages (Redis, CouchDB, MariaDB).

Le plus gros défi technique était de parvenir à quitter l'étape de la preuve de concept, où tout fonctionne comme prévu quand un seul utilisateur lance la génération d'un seul guide à la fois, pour parvenir à un système complet capable de grandir au même rythme que sa fréquentation quand son usage s'accélère. Sans surprise, la virtualisation et des serveurs multiples permet beaucoup de choses de ce côté là...

D'ailleurs, la vitesse a été un point sur lequel nous avons beaucoup insisté. De plusieurs dizaines de minutes pour générer un guide, nous sommes parvenus à passer sous la barre des quelques minutes seulement, parfois quelques dizaines de secondes, selon la complexité des data analysées. En effet, si nous souhaitons aider le rédacteur à rédiger plus vite, autant éviter de le faire patienter plusieurs jours pour obtenir son guide, n'est-ce pas ?

C'est pour cela que le site web est optimisé pour l'affichage rapide, en http/2, et affiche un score Google PageSpeed qui remplit toujours de joie béate quand on l'aperçoit. :-)

Comment utiliser YourTextGuru, concrètement ?

Nous avons beaucoup tourné autour du pot jusqu'à présent dans cet article, mais au delà des bonnes intentions, des algorithmes, de la partie technique ou de la petite histoire, vous avez certainement envie de mieux cerner comment utiliser l'outil YourTextGuru. L'idée était avant tout qu'il soit facile à prendre en main, sans avoir nécessairement à comprendre tout ce qui se trame en arrière-guichet.

Le processus de génération d'un guide, est simple :

1. Il faut d'abord s'inscrire au site (c'est gratuit ; c'est la commande de guides qui est payante seulement, à l'usage, via des packs de jetons). Les nouveaux inscrits profitent en ce moment de quelques jetons offerts pour leur permettre de tester le service.

2. On se rend sur la page de commande de guides.

3. On entre la requête pour laquelle on souhaite un guide. On parle ici de requête, car c'est à partir d'une requête dans un moteur de recherche que nous allons pouvoir constituer le dataset de référence pour le futur guide.

Commander des guides de rédaction

Votre requête

Exprimez votre requête ou votre sujet sous une forme intentionnelle ("comment améliorer sa mémoire grâce à la visualisation" plutôt que simplement "meilleure mémoire"). La qualité du guide dépend directement de la précision de votre requête : plus vous êtes précis, plus le guide est bon.

Nombre de guides

Un guide permet de vous guider pour écrire un texte. Si vous souhaitez avoir plus de liberté et de diversité quant au choix des termes de vos futurs articles, commandez davantage de guides pour une même requête. Chaque guide est unique.

Un guide de rédaction coute 1 jeton

Fig.3. L'interface de saisie de votre demande.

Dans les faits, il est souvent plus facile pour un utilisateur néophyte d'entrer le sujet de son article final (« la pêche à la morue à terre-neuve »), voire le titre (« fin de la pêche à la morue à terre-neuve en 1922 »), ou même l'angle de l'article (« qu'est-ce qui a provoqué la fin de la pêche à la morue à terre-neuve en 1922 ? ») plutôt de des mots clés les uns derrière les autres, sous la forme d'une requête usuelle. En effet, plus la requête est précise, plus le dataset l'est, et plus il est facile d'appréhender le guide généré par YourTextGuru : l'utilisateur non averti y verra des termes plus « logiques », plus faciles à appréhender que des mots qui pourraient n'avoir rien à faire là d'un point de vue analytique « humain », mais qui ont tout de même une forte importance pour un algorithme de moteur de recherche qui les verrait comme des mots ayant tout à fait leur place à cet endroit.

4. Après quelques minutes, le bouton de téléchargement apparaît. Le guide est disponible en deux versions, PDF ou TXT. Le fichier PDF a la particularité de proposer des liens permettant de mieux comprendre qui se cache derrière les entités nommées que l'on peut trouver dans le guide.

Un exemple de guide de rédaction

Pour l'exemple, nous avons généré un guide de rédaction concernant Amith Singhal, et plus particulièrement son récent départ de Google. Je souhaite écrire un article sur le sujet, en me faisant aider pour gagner du temps en évitant les recherches trop poussées et en étant certain d'utiliser assez de termes importants, mais pas trop (je vais pouvoir épater

le service SEO de ma société : le texte va être optimisé du premier coup :-). Je saisis donc la requête « *amith singhal quitte google* » dans YourTextGuru, soit le titre que je vais utiliser pour mon futur article.

Le guide, version texte - ci-dessous un extrait -, se présente ainsi (rappelons que chaque guide est unique, en commandant un second guide pour la même requête, celui-ci sera différent) :

TITRE

amith singhal quitte google

CHAPO

amith singhal quitte google + intelligence + vie + changement

PARAGRAPHE

*amith singhal quitte google + artificielle + philanthropiques + web
pensante + systeme + firme mountain
ingenieur + milliard + recherches
amith singhal quitte google + loyaux services + page sergey
vice president + what journey + google moteur
recherches + changement + technologie*

Le guide indique quels termes il faut utiliser, combiner, pour parvenir à écrire un texte naturel. C'est un canevas qui permet d'avoir une base solide sur laquelle se baser tout en évitant les écueils de la suroptimisation. En se basant sur le guide et en ajoutant les connaissances et l'expertise du rédacteur, plus ou moins connaisseur de la thématique, on parvient à écrire vite et bien. Voici le résultat (les mots indiqués en gras sont les termes proposés par le guide, ils ne seront pas en gras dans l'article) :

TITRE

amith singhal quitte google

Le titre sera donc : **Amith Singhal quitte Google**

CHAPO

amith singhal quitte google + intelligence + vie + changement

Alors que la firme de Mountain View s'apprête à se lancer, avec RankBrain, dans l'**intelligence** artificielle, **Amith Singhal quitte Google** et annonce du **changement** dans sa **vie**.

PARAGRAPHE

amith singhal quitte google + artificielle + philanthropiques + web

Amith Singhal quitte Google alors que l'**intelligence artificielle** arrive dans le moteur de recherche, pour se consacrer à des oeuvres **philanthropiques** hors du **web**.

pensante + systeme + firme mountain

La **firme** de **Mountain View** doit à cette tête **pensante** tout son **système** informatique de recherche.

ingenieur + milliard + recherches

L'**ingénieur** est responsable du succès du moteur qui prend en compte plus d'un **milliard** de **recherche** par mois.

amith singhal quitte google + loyaux services + page sergey

Après de bons et **loyaux services** auprès de **Sergey Brin** et **Larry Page**, **Amith Singhal** quitte **Google**.

vice president + what journey + google moteur

Le senior **vice president** a annoncé son départ du **moteur** de recherche **Google** en twittant « **What a Journey!** ».

recherches + changement + technologie

Les **recherches** du moteur sont maintenant vouées au **changement**, avec une **technologie** qui évolue.

Simple, non ? :-)

Utiliser YourTextGuru pour la recherche documentaire

On a vu que parmi les mots que l'on trouve dans les guides, certains sont particulièrement porteurs de sens par rapport à la requête considérée. En effet, ces mots apportent une certaine expertise au texte.

De fait, il devient utile de défricher de nouveaux sujets en utilisant plusieurs guides comme base de recherche documentaire. En lisant 3, 4, 10 guides différents pour une même requête, on peut rapidement savoir de quoi il s'agit, appréhender plusieurs pistes pour aller plus loin pour creuser le sujet. Un véritable gain de temps.

Utiliser YourTextGuru pour le référencement web

La lettre R&R parle avant tout de moteurs de recherche et de référencement web. Il semble donc légitime de se poser la question de l'efficacité de YourTextGuru pour écrire des textes aptes à bien se positionner dans les résultats de Google.

Bien sûr, le modèle génératif permet de s'affranchir du risque de suroptimisation, et même de s'assurer qu'un texte sera naturel (comparé à ceux que l'on trouve en bonnes places dans les SERP), donc on peut imaginer que d'une façon ou d'une autre il y ait une influence pour le moteur de recherche.

Cependant, nous n'avons pas encore eu l'occasion de faire d'étude quantitatives, car le service est très récent et il n'existe pas encore sur le Web assez de textes écrits avec son aide. Cependant, les premiers retours utilisateurs sont très encourageants (certains sont visibles sur Twitter). Nous avons déjà plusieurs témoignages très positifs sur le gain en temps de rédaction, mais aussi sur le positionnement. De là à généraliser, ce n'est pas quelque chose que l'on peut encore faire. Mais nous travaillons dur pour que l'outil soit le meilleur possible jour après jour...

Cependant, étant donné le prix très peu onéreux du service et les jetons offerts lors de l'inscription, il serait dommage de ne pas tester par vous-même et de vous faire votre propre idée. Alors c'est ici : <https://yourtext.guru> !



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/>