



Améliorez la visibilité de votre site sur Google !

Par Nassoub



Sommaire

Sommaire	2
Partager	2
Améliorez la visibilité de votre site sur Google !	4
Formez-vous au référencement Google !	4
Partie 1 : Introduction au référencement	5
Parlons de référencement...	6
Qu'est-ce que le référencement ?	6
Un peu de vocabulaire	6
Exemple de référencement	6
Mon site est-il référencé sur Google ?	7
Les clés du référencement	8
La région du Triangle d'Or	8
Vous devez être bien positionné	9
Pourquoi ne parle t-on que de Google ?	10
L'avant et l'après-Google	10
L'avant-Google - 1993 à 1998	10
L'après-Google	11
Google c'est aussi...	12
En résumé	13
Bien se préparer pour bien débiter	13
Quel navigateur web choisir ?	14
Le navigateur web	14
Une suite d'extensions à adopter !	15
Les extensions utiles au référencement de Firefox !	15
Ce qui se fait du côté de Google Chrome !	15
En résumé	16
Les robots sur le web	17
Introduction aux robots	17
Nous ne sommes pas seuls	17
Comment fonctionne Google ?	17
Comment Google fonctionne t-il ?	17
Le fonctionnement de Googlebot	17
Quand est-ce que passe le Googlebot ?	18
Le référencement se construit sur la durée	18
Les sanctions de Google	19
En résumé	19
Partie 2 : Optimisez vos pages web pour le référencement	20
Concepts de base d'un site « SEO-Friendly »	20
Un site web, du code HTML	20
Le langage HTML	20
Une question d'ergonomie	22
Ergonomie de base d'un site	22
La règle d'or des 600px : ergonomie et référencement.	23
Pensez à votre visiteur !	25
Performances, performances, performances !	25
Accélérez vos temps de chargement !	25
Google et l'empathie	27
En résumé	27
Communiquez avec Googlebot !	27
Aidez Googlebot à générer le résultat de recherche	28
Google génère ses résultats à partir des pages web	28
Les meta description et keywords ne sont pas encore morts !	29
Deux balises <meta /> qui survivent aux années	29
META description	30
META keywords	30
Question utilité ?	31
Interdire l'indexation d'une page avec le meta robots !	31
Un nouveau meta : Robots	31
Gérer le suivi des liens au cas par cas	32
Interdire l'indexation d'un site avec le robots.txt !	33
Gérez les accès des robots avec robots.txt	33
Bonne pratique	34
En résumé :	35
Un site web, une arborescence, des URL ! (1/2)	35
La base de l'URL : le nom de domaine !	36
Les URL de votre sites	36
Construire des URL "SEO-Friendly"	37
Points fondamentaux de la construction d'URL !	37
Une autre règle d'or : un changement = une redirection.	39
Précisions sur les redirections	40
Problématique des URL dynamiques	41
Un problème d'URL	41
La solution	42
En résumé	42

Un site web, une arborescence, des URL ! (2/2)	42
Le Link Juice ou Jus de lien !	43
Le Google Juice ! (ou Link Juice)	43
Acheminer le jus à travers son site	46
Transmettre le jus aux autres pages	46
Ne gaspillez pas bêtement votre jus !	48
Attention aux 404 !	48
Précisions sur le Google Juice	50
Quel jus est transmis ?	50
Cas particulier	52
En résumé	53
L'importance des mots-clés !	53
Une question de mots-clés !	54
Un mot-clé, des mots-clés	54
Établir une liste de mots-clés pour son site	54
La création de votre site	54
Le suivi des mots-clés	55
Choisissez vos mots-clés grâce à Google !	55
Utilisez Google Insights	55
Le concept de la longue traîne	57
Expliquons la longue traîne	57
En résumé	58
La sémantique et le référencement	59
La sémantique et le référencement	59
Le pari de la sémantique	59
Le référencement à travers l'évolution du web	60
Le référencement évolue à travers le web	60
HTML5 et les micro-données de l'avenir	61
Microformats ou micro-données ?	61
Balisez sémantiquement votre contenu	61
En résumé:	63
Les pièges à éviter et les bonnes habitudes à prendre	64
Les mauvaises pratiques	64
Le « Duplicate Content »	64
Votre contenu est votre force	65
Flash et le référencement	65
Flash et le référencement	65
En résumé	66
Partie 3 : Annexes et autres ressources	66
Découvrez Lynx Browser !	67
Lynx : un navigateur !	67
Le lecteur d'écran Lynx	67
Votre première visite sous Lynx Browser !	69
La navigation sous Lynx Browser	69
En résumé	71



Améliorez la visibilité de votre site sur Google !



Le tutoriel que vous êtes en train de lire est en **bêta-test**. Son auteur souhaite que vous lui fassiez part de vos **commentaires** pour l'aider à l'améliorer avant sa publication officielle. Notez que le contenu n'a pas été validé par l'équipe éditoriale du Site du Zéro.

Par



Nassoub

Mise à jour : 17/10/2012

Difficulté : Facile  Durée d'étude : 2 mois



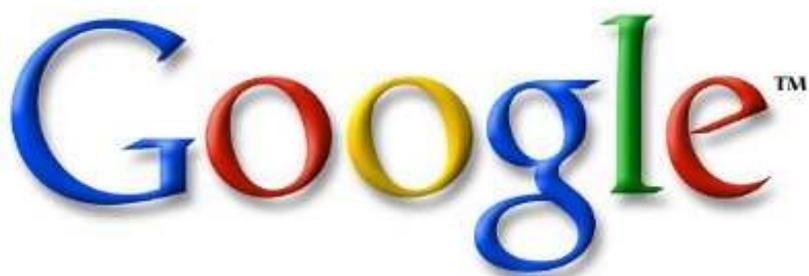
1 160 visites depuis 7 jours, classé 112/792

Salut à vous, mes chers Zéros !

S'il est bien un sujet qui a fait couler beaucoup d'encre chez nous autres Webmasters, c'est bien celui du trafic que notre site génère. Forcément, la finalité d'un site internet c'est bien d'être visité ; on ne crée pas un site web juste pour faire joli. Une question va alors très vite se poser : comment générer ce trafic et comment amener du monde à visiter son site web ? La réponse qui reviendra le plus souvent, c'est celle du passage par la case « amélioration » de votre visibilité sur les moteurs de recherche.

Formez-vous au référencement Google !

Il est vrai qu'au gré de nos recherches sur le web on trouvera beaucoup de documentation sur les technologies qui nous servent à créer des sites telles que HTML5 et CSS3. Le Site du Zéro en est d'ailleurs un excellent exemple. Mais il est aussi vrai que l'on trouve moins de documentation sur la manière de faire vivre son contenu, c'est-à-dire faire en sorte qu'il soit vu et lu. C'est à cette problématique que nous allons tenter de répondre par le biais de ce tutoriel.



Google ! Le moteur de recherche dont on va

beaucoup parler !

Mon rôle à moi va être de vous fournir de manière claire et compréhensible les bases du **référencement** sur Google, le **Roi** des moteurs de recherche. Le référencement regroupe l'ensemble des stratégies qui vous serviront à amener du monde sur votre site. Concrètement, nous allons apprendre dans ce tutoriel à construire de meilleurs sites web, à en améliorer la réputation, et surtout, faire en sorte que votre site web ressorte dans les premiers résultats de recherche sur Google. Votre site générera ainsi naturellement du trafic et prendra vie !

A la fin de ce cours, vous saurez ce que sont les robots sur le web, vous saurez faire indexer certaines pages de votre site mais pas d'autres, vous saurez également créer des mises en page ergonomique. Nous parlerons aussi de maillage interne, de sémantique, de micro-données ou de Google Juice... bref ! Vous maîtriserez ces notions dont vous n'avez certainement jamais entendu parler, mais qui sont pourtant cruciales pour attirer des visiteurs sur son site web.

Pour prétendre à travailler le référencement d'un site web, il est logique de savoir le construire. Il est donc important d'avoir un niveau basique en HTML5 et CSS3. Pas besoin d'être un expert, il vous faut seulement de bonnes bases dans ces

standards.(*mise à niveau HTML5/CSS3*)

Partie 1 : Introduction au référencement

Cette première partie purement introductive vous permettra de faire un point et de vous équiper correctement avant de commencer à apprendre réellement le référencement dans la seconde partie. Voyez ce début du cours comme un complément du tutoriel HTML5/CSS3 du Site du Zéro. Vous avez appris à créer un site ? Super ! Maintenant, il faut le faire vivre !

Parlons de référencement...

Je suis content de voir que vous m'avez suivi dans cette aventure. On va donc pouvoir commencer à parler de référencement de manière sérieuse : savez-vous ce dont il s'agit ? Dans ce chapitre, nous n'allons pas commencer à toucher aux codes sources qui composent votre site, mais plutôt mettre en place les fondations du référencement dans votre tête.

Qu'est-ce que le référencement ?

Un peu de vocabulaire

Du référencement, du positionnement ou du SEO ?

C'est la question **la plus importante** à laquelle je dois répondre dans ce cours : qu'est-ce que le référencement. Étant donné que ce mot risque de revenir assez souvent, autant le définir tout de suite.

N'importe quel expert le ferait remarquer : « référencement » n'est pas le terme exact pour désigner ce que l'on va étudier tout au long de ce cours. « Référencer » sur le web, c'est faire en sorte qu'un site web soit présent dans un index. Dans le cas de Google, il suffit donc que votre site soit présent dans les résultats de recherche pour qu'il soit considéré comme « référencé ».

Mais vous, vous ne voulez pas seulement à être présent, vous voulez avoir aussi une bonne place dans ces résultats pour que votre site attire un max de visiteurs. Ainsi le meilleur terme pour désigner l'ensemble des actions que l'on entreprend pour que son site bénéficie de la meilleure place dans les résultats de recherche serait plutôt « positionnement ».



Mais pourquoi utiliser le terme « référencement » alors ?

C'est simplement parce que le mot « référencement » s'est vite imposé par abus de langage comme un terme générique pour désigner « référencement » et « positionnement ». Pour parler aussi bien de la présence de son site dans l'index que de sa position, on a vite commencé à n'utiliser que le mot « référencement ». N'hésitez donc pas à parler du référencement de votre site sur des forums ou dans une conversation, c'est le mot qui sera le plus compris.



Le terme anglais est très précis. Les Anglais parlent de « Search Engine Optimisation » : le SEO. Cette expression est aussi très utilisée en France.

Exemple de référencement

Ce que nous souhaitons obtenir pour votre site

Ce point de vocabulaire passé, voici ce que l'on va souhaiter obtenir pour votre site web. Prenons l'exemple de ce vidéoclub.



Une métaphore de ce qu'il se passe sur le web.

Le référencement

Les films présents sur la droite dans ce vidéoclub sont « référencés » ; ils sont présents dans l'index du vidéoclub. Si vous vous rendez au guichet et que vous demandez l'un de ces films, on pourra vérifier rapidement sur un ordinateur que le film que vous souhaitez est présent et dans quel rayon il se trouve. Ceci est le référencement.

Le positionnement

Si vous regardez la même image, vous devriez voir qu'un film est mis en avant sur la gauche. Il est séparé des autres films, et en plus, il bénéficie d'une petite affiche qui retiendra probablement votre regard lorsque vous allez passer dans ce rayon. Ceci est le positionnement, par abus de langage on peut aussi dire qu'il est « bien référencé », comprenez « bien positionné ».

Ce que nous souhaitons obtenir pour votre site, c'est exactement la même chose que ce film qui est sur cette affiche : qu'il soit mis en valeur et visible, mais sur Google. Être bien référencé sur Google signifie être dans les premiers sites sur la page de résultat, vous verrez tout au long de ce cours que ce n'est pas une mince affaire, il va falloir peut-être réapprendre totalement à concevoir vos sites web, mais ne vous en faites pas, on est là pour apprendre tout cela. 😊

Mon site est-il référencé sur Google ?

Faites le test !

Ce test est tout à fait possible. Vous pouvez parfaitement demander au moteur de recherche si votre site est référencé, et ce depuis la barre de recherche Google. Pour cela c'est très simple. Il existe un mot-clé `site :` utilisable dans la barre de recherche de Google. Vous le ferez suivre par le nom de domaine du site dont vous voulez tester l'existence dans l'index.



La Google Search Bar possède aussi ses petits secrets.

Par exemple, si je souhaite vérifier si le Site du Zéro est référencé sur Google, j'ouvre la [page d'accueil de Google](#), et je tape la recherche qui suit dans le moteur de recherche.

Code : Autre

```
site:siteduzero.com
```

Étant donné que le Site du Zéro existe dans l'index de Google, le moteur de recherche me retourne toutes les pages du Site du Zéro qu'il possède. En revanche, si je refais le même test avec un site n'existant pas dans l'index de Google, par exemple :

Code : Autre

```
site:aucunresultat.com
```

Google indique bien qu'il ne connaît pas ce site, celui-ci n'est donc pas référencé. N'allez pas réserver le nom de domaine aucunresultat.com exprès pour me faire mentir. 😊



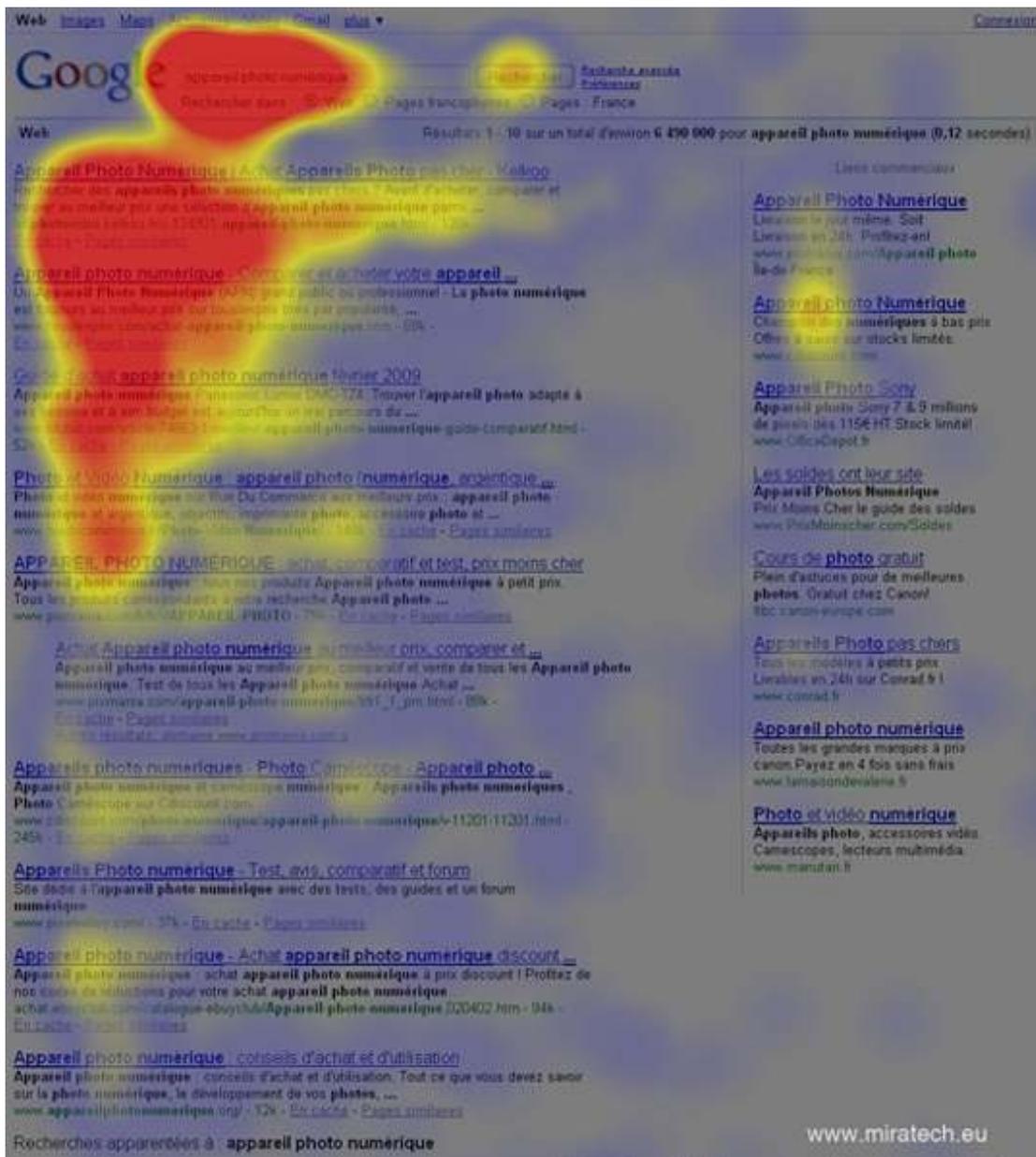
Notez bien que vous ne devez pas renseigner le protocole HTTP (*http://*), Google n'en a pas besoin, contentez-vous de lui indiquer le nom de domaine précédé de `site:` et il saura travailler tout seul.

Les clés du référencement

La région du Triangle d'Or

Soyez dans le top 3 !

L'expression « Triangle d'Or » désigne sur le web la région chaude de la page de résultat de Google. C'est sur cette région que le regard de l'internaute se posera en priorité absolue, et c'est donc ces sites qu'il visitera. Après avoir tapé une recherche, Google affiche une page présentant les résultats sous forme de liens vers différents sites. Plus votre site sera placé dans les premiers résultats, plus il sera visité.



Miratech sur la SERP (Search Engine Results Page) de Google !

Ici, la recherche est menée sur la SERP de Google. Vous pouvez voir que le regard de l'internaute se posera en priorité sur les 3 premiers résultats de recherche. Cela forme une région triangulaire que l'on appelle le « Triangle d'Or ». Cette étude a un but précis : mettre en évidence le comportement d'un internaute sur Google. Vous devrez être dans les premiers résultats de recherche pour que votre site soit visité, car sur le web, tout va très vite.

Pour l'internaute, sa recherche doit se faire vite, et même **très** vite. L'internaute ne perd pas son temps sur Google, il veut une information rapidement et c'est pourquoi il clique sur les premiers liens qui s'affichent. C'est exactement ce que **Miratech** démontre avec le Triangle d'Or. Lorsque vous êtes indexé sur Google, vous êtes certes présent sur les serveurs de la firme mais si votre référencement n'est pas travaillé, vous ne serez pas présent dans les premiers résultats Google. De ce fait vous ne serez pas visité.



Petite introspection. Vous, lorsque vous recherchez quelque chose sur Google, cela vous arrive souvent de vous aventurer sur les pages à deux chiffres ou de passer simplement aux autres pages de résultat ? 🤔

Vous devez être bien positionné

Travaillez votre référencement

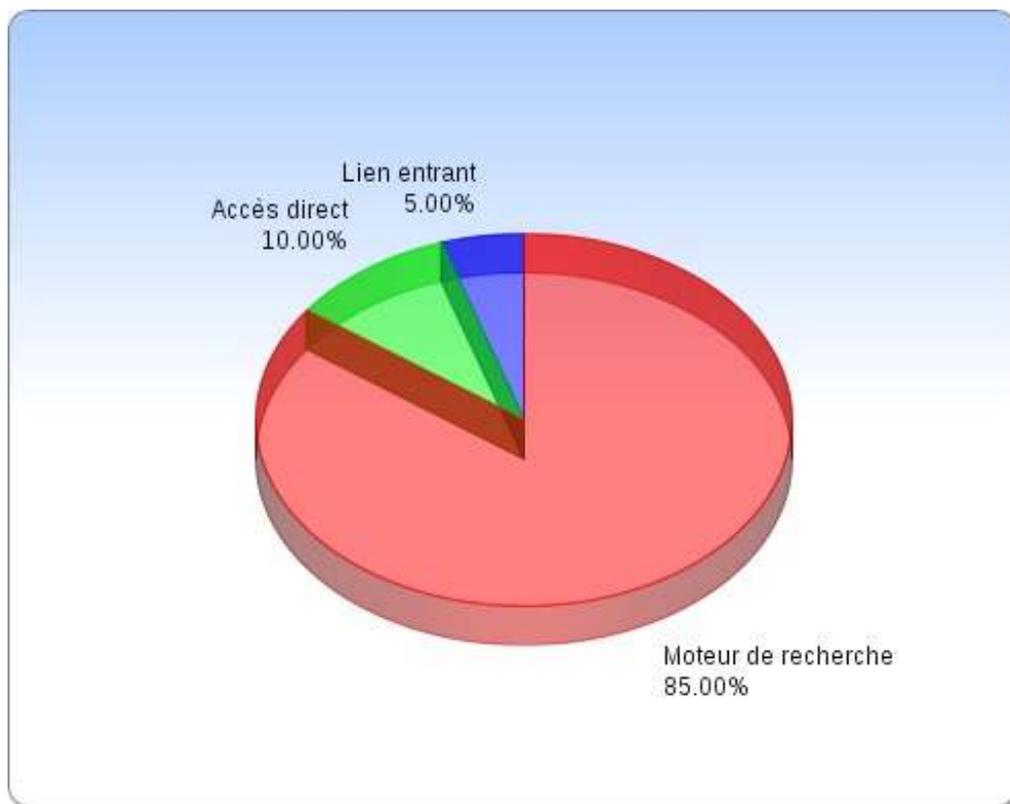
Pour que votre site soit bien positionné (et puisse vivre), il faudra batailler ferme sur le référencement, il n'y a pas de miracle. Si vous êtes dans les abîmes des résultats de Google **personne ne visitera votre site**. Ce qui revient à dire que si votre site n'est pas

dans le Triangle d'Or : personne ne profitera de vos pages.

Ce Triangle d'Or est également délimité par une ligne imaginaire horizontale située à 600px. Si vous êtes sur la première page des résultats Google mais que vous vous trouvez sous cette ligne, c'est-à-dire sous le 6e résultat, vous avez là encore peu de chance d'être visité. Miratech a mis en évidence dans une autre étude qu'en dessous du 5e résultat, l'utilisateur préfère retaper une nouvelle recherche plutôt que de s'aventurer plus loin...💡

Quelques chiffres

D'une manière générale, un site web génère son trafic grâce aux moteurs de recherche. En moyenne, 85% de ses visiteurs viennent des moteurs de recherche, Google en tête. Pour le reste, 10% accèdent au site soit par accès direct (favori, barre d'adresse) et 5% par un lien entrant trouvé ailleurs sur le web.



J'espère vous avoir fait comprendre que vous n'avez vraiment pas le choix : il va falloir travailler à fond votre référencement et arriver à entrer dans le Triangle d'Or ! Sans quoi votre site ne générera jamais le trafic dont vous rêvez.

Pourquoi ne parle t-on que de Google ?

Google au centre des débats

Depuis tout à l'heure, nous ne parlons que de Google, mais pourquoi ? Vous êtes peut-être fan de Yahoo! et je suppose que vous aimeriez en priorité que votre site soit bien référencé sur ce moteur... vous pensez donc que ce cours qui parle surtout de Google ne va rien vous apporter. Et bien ne pensez surtout pas cela !

Pour chasser vos inquiétudes à ce sujet, je dois vous parler de l'histoire de Google et plus généralement du web des années 1990 au web d'aujourd'hui.

L'avant et l'après-Google

Le web est immense et depuis sa création, des moteurs de recherche se sont donné pour objectif de permettre aux internautes de trouver l'information rapidement dans cette masse de contenu. Dans [un article de l'Euromag](#), Ryan Franck propose très justement une division de l'histoire du référencement en deux périodes : l'**avant-Google** et l'**après-Google**. Je vais également suivre cette idée et diviser cette présentation historique avec ces deux parties, vous comprendrez mieux pourquoi nous allons beaucoup parler de Google.

L'avant-Google - 1993 à 1998

Le référencement simple

À cette époque, Google n'existait pas encore, mais la science du référencement existait déjà. Elle était animée par un leader à l'époque incontesté, nommé *Altavista*. Il s'agissait d'un moteur de recherche qui s'appuyait sur un système « simple » pour déterminer ses résultats de recherche : les **mots-clés**. En d'autres termes, vous assurer un bon placement dans la page de résultats était simple : tout ce que vous aviez à faire était de renseigner les bons mots-clés et en bon nombre dans les balises `<meta />` de vos pages web. Dès lors, une comparaison était effectuée par le moteur de recherche entre les mots-clés tapés dans la barre de recherche, et les balises `<meta />` des pages web indexées. Avoir un bon référencement était donc très simple.



Rappel : Les balises `<meta />` sont les balises se trouvant dans la section `<head>` de votre code HTML. Le contenu de ces balises sert à donner des informations sur le contenu même de votre page aux moteurs de recherche, entre autres.

L'algorithme sur lequel le système d'Altavista était fondé ne tenait compte que de cela. Les utilisateurs qui n'adhéraient pas à Altavista pouvaient toujours utiliser des moteurs de recherche tels que Yahoo! ou Lycos qui existaient déjà à cette époque. Encore une fois, et j'attire vraiment votre attention là-dessus, le référencement à cette époque était une science très simple par rapport à aujourd'hui. Les balises `<meta />` suffisaient à s'attirer les grâces des moteurs de recherche. Aujourd'hui, c'est un véritable parcours du combattant. 🐜



Un logo que vous ne pouviez pas manquer il y a encore 15 ans.



Notez qu'avant Altavista (avant 1995 donc), les moteurs de recherche étaient encore plus simples. Ils se basaient sur le nombre d'occurrences d'un mot-clé dans une page pour positionner les sites.

C'est en 1998 que cette situation a été bouleversée par un nouvel arrivant : **Google** !

L'après-Google

Le référencement devient très complexe

En 1995, deux étudiants de l'université de Stanford, Larry Page et Sergeï Brin, font équipe sur un projet qu'ils baptisent « Backrub ». L'idée des deux hommes est de créer un nouveau moteur de recherche plus performant qu'Altavista. Ils parviennent à réaliser un nouvel algorithme révolutionnaire, capable de déterminer la pertinence d'un site en fonction du nombre de liens entrants vers celui-ci et de leur qualité. Tout a changé à ce moment-là : le bourrage de mots-clés dans les meta du site ne suffisait plus, il fallait maintenant que d'autres sites fassent des liens vers le nôtre pour qu'il soit reconnu comme pertinent, et donc bien référencé.

Google était né (il prendra son nom définitif 3 ans plus tard, en 1998).



Sergeï Brin et Larry Page, les pères fondateurs de

Google toujours à la direction de l'entreprise aujourd'hui.

Tout de suite, c'est l'explosion. Ce système révolutionnaire réunit un nombre d'utilisateurs colossal, et détrôna sans grand mal Altavista.



Altavista ayant sombré face à Google, le moteur de recherche a été racheté par une entreprise nommée Overture, puis par Yahoo!.

Vous l'aurez compris, ce qui a fait le succès de Google, **c'est son algorithme**. L'ère des meta est révolue avec la chute d'Altavista et nous sommes entrés à ce moment-là dans l'ère du **contenu**, qui est depuis l'élément clé de chaque page web. Depuis, Google n'a jamais cessé sa croissance exponentielle, et est aujourd'hui en situation de quasi-monopole sur le marché des moteurs de recherche.



Ne pensez pas que faire beaucoup de liens entrants suffira à vous garantir un bon référencement. Cela serait beaucoup trop facile... et ce tutoriel n'aurait pas lieu d'exister. 🤔

En effet, si Google a su trouver la clé de son succès, celle-ci n'a pas cessé de s'enrichir et de se complexifier au fil du temps. Depuis, de nouveaux algorithmes ont été créés et ajoutés à Google. Ceux-ci tiennent compte de la qualité de votre code HTML, de votre présence sur les réseaux sociaux et de beaucoup d'autres éléments. Le référencement n'a donc pas cessé de se complexifier et créer des liens ne suffit plus : Google se base aujourd'hui sur plus de 200 critères de pertinence pour déterminer le positionnement des sites dans son index. C'est donc sur plus de 200 critères qu'il va falloir travailler pour se démarquer des autres sites.



Chez Google, sachez par ailleurs qu'un département de développement appelé « Search Quality » est composé de développeurs et d'ingénieurs chargés uniquement de mener des recherches et d'améliorer constamment l'algorithme d'indexation. C'est vous dire que les choses bougent toujours, et elles n'ont jamais été aussi complexes qu'aujourd'hui !

Vous aurez donc tout à gagner à vous concentrer essentiellement sur Google car celui-ci tient littéralement le marché des moteurs de recherche dans sa main, tant il est en situation de monopole. Un très léger pourcentage des internautes se partage d'autres moteurs de recherche tels que Bing et Yahoo!, mais lorsqu'on sait que leurs systèmes sont très inspirés de ceux de Google, pratiquement aucune conversion ne sera nécessaire.

Ainsi, même si ce tutoriel portera essentiellement sur Google, ce que vous apprendrez sera réutilisable sur une très grande majorité des moteurs de recherche, vous ne négligerez ainsi aucune source de trafic pour votre site. 😊

Google c'est aussi...

... une foule d'autres services qui ont de près ou de loin attiré au SEO

Google vous fournira également **les meilleurs outils** pour votre site web. [Google Analytics](#) vous permettra d'analyser toutes les sources de trafic et autres statistiques de votre site, tandis que les [Google Webmaster Tools](#) vous aideront à débusquer vos erreurs et à voir votre site comme Google le voit, [Google AdSense](#) et [Google AdWords](#) vous permettront également de passer de l'autre côté de la barre et d'aborder le référencement payant.



Et tout ça, nous le verrons ensemble dans ce cours, en même temps que nous aborderons toutes les facettes du référencement de votre site web !



Au fait, si vous n'avez pas de compte Google, je vous conseille de vous en créer un tout de suite. Cela vous fera gagner un peu de temps, et un Webmaster, un vrai, se doit d'avoir un compte Google.

Les règles volatiles que Google ne communique pas

La dernière chose à savoir vis à vis du référencement est le fait que Google ne communique pas sur ses critères de pertinence. À part le département « Search Quality » évoqué plus haut, personne ne sait réellement comment Google évalue votre contenu.



Mais alors ? C'est quoi ce tutoriel ? Tu bosses pour Google ?

Pas vraiment, et moi-même, je ne connais pas ces critères de manière exacte. Le fait est que si Google ne communique pas, des centaines de milliers de tests ont été effectués ces dernières années sur des points particuliers pour comparer la manière dont le moteur positionne les contenus. À partir des résultats de ces tests, on est capable de connaître assez précisément le comportement de Google, et de là, on sait ce qu'il faut faire pour qu'un site soit bien référencé. On sait donc aujourd'hui comment créer des contenus « *SEO-Friendly* ». Le but de ce cours va être de vous initier à ces techniques de référencement : vous verrez que l'on va avoir beaucoup à faire !

En résumé

- Le référencement consiste en l'optimisation de son site pour obtenir une meilleure place dans les moteurs de recherche.
- Votre trafic et la survie de votre site dépendent de votre place dans les résultats de la SERP de Google.
- Nous nous concentrerons dans ce cours sur Google, car il monopolise le marché des moteurs de recherche.
- Google vous propose une suite d'outils qui vous aideront dans votre tâche d'optimisation : créez un compte Google !

Bien se préparer pour bien débiter

Pour bien débiter dans le monde du référencement, il va falloir bien s'équiper. Se préparer pour se former au référencement va passer par quelques points relativement simples. De quels navigateurs s'équiper et pourquoi ? Quels sont les plug-ins incontournables ? C'est maintenant que l'on va se pencher sur ces questions !

Quel navigateur web choisir ?

Le programme essentiel qui va nous servir à travailler notre référencement est bien évidemment le navigateur web. Il existe pourtant une multitude de navigateurs, et tous ont leurs points forts et leurs points faibles. Nous allons voir quels sont les navigateurs qui vont vraiment nous être utiles dans notre dure tâche de référencement. 🤖

Le navigateur web

Quel est le meilleur navigateur web pour le SEO ?

Disons qu'il n'y a pas vraiment de meilleur navigateur mais qu'il y a plutôt plusieurs navigateurs qui sont excellents. À vrai dire, tout navigateur peut être très bien pour le référencement : on n'a pas vraiment besoin d'un navigateur qui sache faire le café pour le coup ! On ne demande qu'une seule chose au navigateur : qu'il soit à jour au niveau des standards. Un navigateur capable d'afficher correctement les technologies qui font le web sera un bon navigateur pour référencer vos sites.



Par ailleurs, lorsque vous avez appris HTML et CSS, on vous a dit de toujours utiliser plusieurs navigateurs pour tester la compatibilité : ce conseil vaut toujours.

Voici une liste des principaux navigateurs web d'aujourd'hui. Vous pouvez les utiliser pour suivre ce tutoriel, et pour travailler le référencement de votre site. Tous sont en théorie gratuits à part Internet Explorer qui nécessite l'achat d'une licence Windows.

Icône	Nom	OS	Téléchargement	Description
	Internet Explorer		Fourni par défaut avec Windows.	Un navigateur web signé Microsoft et intégré par défaut à toutes les versions du système d'exploitation Windows. Si vous tournez actuellement sous Windows, alors vous disposez de ce navigateur. Je ne vous recommande pas de l'utiliser, car les versions les plus récentes ne sont pas disponibles pour les systèmes comme Windows XP qui sont pourtant massivement utilisées encore aujourd'hui. En revanche si vous utilisez Windows 7, vous pouvez bénéficier des dernières versions d'Internet Explorer qui sont performantes.
	Mozilla Firefox	 	Téléchargement	Un navigateur web gratuit et extrêmement performant signé par la Mozilla Foundation. Je ne saurais que vous le recommander !
	Google Chrome	 	Téléchargement	Un autre navigateur web signé cette fois-ci par Google. Le programme est basé sur un projet open-source baptisé Chromium, mais il connaît quelques controverses vis à vis du respect des informations personnelles de ses utilisateurs. Il est néanmoins recommandé de l'utiliser tant il est performant.
	Chromium	 	Téléchargement	Chromium est le projet open-source sur lequel est basé Google Chrome. Il s'agit en fait de Google Chrome lavé des ajouts de Google. Vous pouvez l'utiliser si vous êtes inquiet au sujet des questions du respect des informations personnelles.
	Opera	 	Téléchargement	Opera est un navigateur web de très grande qualité proposé gratuitement par l'entreprise Opera Software. Il est moins utilisé que les 4 précédents mais demeure un excellent produit.
	Safari	 	Téléchargement	Safari est le navigateur web d'Apple qui équipe par défaut tous les ordinateurs Apple. C'est un peu l'Internet Explorer de la pomme de ce point de vue, je n'ai pas pour habitude de l'utiliser, mais il bénéficie d'une très bonne réputation.

Il y a deux navigateurs en particuliers qui se démarquent totalement des autres produits : Mozilla Firefox et Google Chrome/Chromium. En effet, ceux-ci possèdent des communautés d'utilisateurs très grandes, qui comptent de brillants

développeurs. Ces développeurs créent tous les jours de nouveaux « plug-in », dont certains seront incontournables pour nous aider dans notre travail de référencement. Voyons dès maintenant quels sont ces plug-in incontournables pour Firefox et Chrome/Chromium.

i Google Chrome et Chromium sont très similaires. Il s'agit à 95% des mêmes navigateurs. L'un est simplement supporté par des développeurs bénévoles alors que l'autre est supporté par Google. Une extension Google Chrome sera également compatible avec Chromium.

Une suite d'extensions à adopter !

Les extensions utiles au référencement de Firefox !

Améliorez les possibilités de Firefox

La fondation Mozilla a eu la bonne idée de créer un site internet officiel pour regrouper tous les modules complémentaires développés par la communauté pour plus de clarté. Ainsi, vous pourrez trouver toutes les extensions disponibles pour Firefox à cette adresse : [Addons Mozilla](#).

En arrivant sur ce site, vous pouvez apercevoir une barre de recherche intégrée. Vous allez pouvoir l'utiliser si vous savez exactement quel type d'extension vous recherchez. Pour notre part, nous voulons des extensions qui vont nous être utiles au référencement web.

Vous avez de la chance, comme je suis sympa je vais vous faire de suite une liste des extensions que j'utilise depuis un bon moment déjà !



i L'ajout d'une extension à Firefox nécessitera le redémarrage du navigateur.

Attention ! Car plus vous ajouterez d'extensions à votre navigateur, plus celui-ci sera gourmand en mémoire ! La réactivité de votre système en sera donc sensiblement réduite.

Icône	Nom de l'extension	Page Mozilla Addons	Description
	Web Developer	Page officielle	Web Developer est une extension Firefox qui se caractérise par l'ajout d'une toolbar à votre navigateur. Elle permet facilement de désactiver les feuilles de styles, les images, les codes JavaScript... Bientôt, vous ne pourrez plus vous en passer.
	Firebug	Page officielle	Firebug complète Firefox en lui ajoutant des outils de développement web. Il vous sera notamment utile pour repérer vos erreurs dans vos codes Javascript, mais aussi mettre en évidence les blocs qui composent votre page web. Appuyez sur la touche F12 pour ouvrir ces nouveaux outils.
	User Agent Switcher	Page officielle	Du même auteur que Web Developer, cette extension Firefox vous permettra de modifier facilement votre User-Agent. Nous verrons au chapitre suivant ce qu'est un User-Agent et l'utilité de chercher à le modifier en référencement.
	WebRank Toolbar	Page officielle	Une barre d'outils qui vous donne un aperçu complet de ce que vaut le site ouvert dans Firefox en matière de référencement.
	KGen	Page officielle	KGen est une extension Firefox qui sera capable de vous fournir une liste de mots-clés pertinents pour la page web que vous visitez.

Vous devez très certainement vous demander à quoi vont servir ces extensions pour votre navigateur, et c'est normal. Vous prendrez bientôt conscience de l'utilité de ces extensions, et ce dès le prochain chapitre !

Ce qui se fait du côté de Google Chrome !

Des extensions sur le Chrome Web Store

Chez Google, on a procédé comme chez Mozilla et on a regroupé toutes les extensions disponibles pour Chrome dans un site web officiel dédié. Le site web où sont référencées toutes les extensions disponible pour le navigateur Google Chrome est le [Chrome Web Store](#).



Bon point pour Google Chrome par rapport à Firefox : l'ajout d'extensions ne requiert généralement aucun redémarrage du navigateur. Vous allez donc pouvoir utiliser vos nouvelles extensions plus rapidement. Comprenez bien que le fait que vous soyez sous Google Chrome ou Chromium ne change strictement rien. Le Chrome Web Store sera fonctionnel pour Chrome comme pour Chromium.

Bien ! À la manière de ce que l'on a fait pour Firefox un peu plus haut, voyons ici les extensions intéressantes que l'on va pouvoir utiliser avec Google Chrome pour le référencement. Pour Google Chrome, on va avoir besoin d'un peu moins d'extensions que sur Firefox. Par exemple, Firefox a besoin de l'extension Firebug afin de pouvoir gérer plus d'outils de développement, alors que Google Chrome propose des outils similaires en natif. J'en ai quand même retenu deux qui pourront vous servir.



Les extensions Google Chrome ne sont évidemment pas les mêmes que les extensions Firefox.

Nom de l'extension	Page Google Web Store	Description
Page Rank Status	Page officielle	Page Rank Status est une extension Google Chrome qui va vous permettre d'afficher dans le navigateur le Page Rank ainsi que beaucoup d'autres informations comme le rang Alexa de la page que vous êtes en train de visiter.
SEO Site Tools	Page officielle	SEO Site Tools est une extension Google Chrome incontournable qui va vous fournir des informations très poussées sur la page web que vous êtes en train de visiter. Elle va notamment vous permettre de lister les headings (<h1>,<h2>,...) ainsi que beaucoup d'autres choses.

Sans vouloir animer de débats, ma préférence personnelle va à Firefox, car il est considéré depuis vraiment longtemps comme l'ami des développeurs. La somme d'extensions qu'il propose, ses performances ainsi que sa communauté en font vraiment un allié de choix pour toute tâche de développement (au sens large) sur le web. Je ne peux donc que vous conseiller de vous équiper de ce navigateur web au quotidien.

En résumé

- Utilisez un navigateur web à jour.
- Une suite d'extensions pour navigateur web est là pour vous simplifier la vie.
- Utilisez aussi votre éditeur de texte habituel.

Les robots sur le web

Sur le web, nous ne sommes pas seuls. Il y a nous, les humains qui visitons les sites web et les faisons vivre, mais il y a aussi d'autres visiteurs : les robots ! Il en existe des centaines qui arpentent le web tous les jours, et chacun d'eux a une tâche précise. Qui sont-ils et font-ils sur le web ? Apprenons à les connaître. 😊

Ne sautez en aucun cas ce chapitre, sinon vous ne comprendrez vraiment plus rien à la suite de ce tutoriel !

Introduction aux robots

Nous ne sommes pas seuls

Qui sont les robots ?

N'oubliez pas que le Terminator existe, ce n'est absolument pas le propos ici ! Les robots, dans le monde de l'informatique, sont des programmes réalisés par des développeurs afin de réaliser une tâche précise. Vous en avez très probablement côtoyé sans le savoir. Si vous êtes un amateur de jeux vidéo et que vous jouez souvent en solo, il arrive souvent que vous ayez quand même des ennemis à affronter. Seulement, ces ennemis ne sont pas dirigés par des humains qui joueraient en même temps que vous : ce sont des robots !

Sur le Site du Zéro, Zozor pourrait être aussi considéré comme un robot car lorsque vous vous inscrivez au Site du Zéro, vous recevez un message privé de la part de Zozor vous souhaitant la bienvenue. Seulement, le compte de Zozor n'est en réalité dirigé par personne, il est programmé lui-même pour vous envoyer ce message de bienvenue à l'inscription !

Les robots du web !

Je le disais en introduction de ce chapitre, les humains ne sont pas les seuls visiteurs sur le web, il existe donc des robots. En fait, sur le web il existe même des centaines et des centaines de ces robots qui remplissent des tâches diverses et variées sur les sites. Ils ont généralement été créés par des entreprises pour leur propre besoin. Voulez-vous un exemple ? Si vous diffusez de la publicité sur votre site web, c'est un robot qui se rend sur votre site afin de déterminer à l'avance quel type d'annonce est le plus approprié à votre site web en fonction de son contenu.

Googlebot, le robot de Google !

Cette introduction aux robots maintenant finie, parlons plutôt d'un autre robot, qui est certainement le plus connu de tous sur le web, il s'agit de « Googlebot » : sachez que le fonctionnement même de Google dépend complètement de ce robot !

Comment fonctionne Google ?

Comment Google fonctionne t-il ?

Google fonctionne grâce à Googlebot

Si je vous dis que Google est un moteur de recherche, je ne vous apprend rien. Seulement savez-vous réellement comment fonctionne un moteur de recherche ? Le rôle d'un moteur de recherche est de permettre à son utilisateur de trouver rapidement une information sur le web en lui renvoyant une liste des sites les plus pertinents. Cette tâche est complexe, car il existe une multitude de sites. Mais Google doit forcément connaître le contenu de tous ces sites afin de renvoyer le meilleur d'entre eux : alors comment fait-il pour récupérer ces informations ? C'est simple, il passe par Googlebot !

Google n'est pas magicien, il a besoin de quelqu'un pour aller visiter tous les sites du monde et rapporter le contenu de ces sites afin de les évaluer. Vous imaginez bien que ce n'est pas un être humain qui a cette responsabilité, le web c'est aujourd'hui plus de 2 milliards de sites, vous imaginez le boulot... Alors Google a eu l'idée simple de passer par un robot et de le baptiser Googlebot ! Des centaines de milliers de Googlebots sont donc lancés constamment sur les sites web et ils retournent ainsi des informations en permanence à Google, lui permettant de peaufiner son index, et de modifier le positionnement des sites sur la SERP.

À l'heure où vous lisez ces lignes, il y a d'énormes chances pour qu'un Googlebot soit précisément en train de les lire aussi et de les retourner à Google ! Au final, Google fonctionne grâce à ses Googlebots. Un moteur de recherche n'est donc rien de plus qu'un très vaste programme informatique, et il n'y a aucune intervention humaine lors de l'indexation d'un site web.



Le fonctionnement de Googlebot

Un programme en plusieurs étapes

Googlebot a été programmé par Google selon plusieurs critères. Il est en effet très perfectionné et comprendre son fonctionnement est l'une des clés pour bien réussir son référencement. Nous allons donc voir ici le comportement que Googlebot adoptera au moment où il verra votre site web.

1. Googlebot arrive sur votre site web

À son arrivée sur votre site web, Googlebot va commencer par analyser votre page web, en regardant son code source HTML. Il enregistre ce code source, et le fait parvenir à Google.

2. Googlebot explore les liens `<a>`

Googlebot ne va pas s'arrêter là : il va repérer tous vos liens et les explorer, de la même manière qu'un utilisateur cliquerait dessus.

3. Googlebot arrive sur de nouvelles pages web

De la même manière que précédemment, il va récupérer les codes HTML de ces nouvelles pages et les faire parvenir à Google.

4. Googlebot explore à nouveaux les liens

Googlebot va encore explorer tous les liens présents, avant d'enregistrer les codes HTML des pages qu'il atteindra... et ainsi de suite.

Voilà exactement comment Googlebot fonctionne, vous pouvez le comparer à un utilisateur qui ouvrirait tous les liens en enregistrant vos pages HTML, il ne fait rien d'autre. Quelques heures plus tard, Google aura indexé ces pages, et elles seront accessible sur simple recherche Google. Nous verrons bien sûr que plus tard, il sera possible d'imposer des restrictions à Googlebot.



Si Googlebot voit 3500 liens sur vos pages web, il ne fera pas de tri et ouvrira les 3500 liens et indexera donc les 3500 pages visées. Nous verrons qu'il sera possible de demander à Googlebot d'ignorer des liens, des pages, des dossiers ou même des sites complets. Vous pourriez souhaiter que certains de vos contenus n'apparaissent pas sur Google (comme vos pages d'administration, par exemple).

Googlebot ne s'intéresse qu'à votre contenu !

Vous l'avez très probablement deviné : Googlebot ne s'intéresse qu'à votre contenu. Dans les chapitres précédents, j'avais utilisé l'expression « ère du contenu » car nous sommes vraiment dans l'ère où le contenu est l'objet central d'une page web et Googlebot a été développé autour de cette idée. En effet, Googlebot ne voit pas du tout les pages de la même manière qu'un internaute humain : votre navigateur web à vous affiche le contenu, mais le met également en forme. Vous êtes capable de distinguer un menu, des images, une barre de navigation... Googlebot ne fonctionne pas du tout comme cela, il ne voit que la source HTML de votre page web, pas sa mise en forme.

Dans la suite de ce tutoriel, vous aurez tôt ou tard besoin de voir vos pages web de la même manière que Googlebot les voit. Pour cela, il va nous falloir un « lecteur d'écran ». Il s'agit d'une catégorie de navigateur web permettant de n'afficher que le contenu d'un site, **en désactivant CSS, JavaScript et images**. Vous pouvez dores et déjà lire une annexe sur le navigateur **Lynx Browser**, un lecteur d'écran très puissant, vous y apprendrez à lire vos pages à comme le fait un Googlebot !

Quand est-ce que passe le Googlebot ?

Les heures de passage de Googlebot

Le Googlebot ne pourra pas être présent éternellement sur votre site. En effet, il va falloir aussi le libérer de temps en temps et le laisser aller voir ailleurs... seulement, on ne sait pas vraiment à quelles heures Googlebot passe sur un site, tout simplement parce que c'est différent pour chaque site web. Googlebot est capable de déterminer de lui-même quand il va devoir passer sur un site web ou non, et pour cela, il va se baser sur les premiers jours d'existence de votre site. Le jour où vous aurez fini de développer votre site et qu'il sera en ligne, vous allez devoir le signaler à Google pour qu'un Googlebot soit envoyé sur votre site en vue de l'indexer, nous apprendrons à le faire. Google va alors indexer pour la toute première fois votre site web.

Dans les premiers jours, Googlebot va avoir tendance à passer très souvent sur votre site pour vérifier si votre contenu est mis à jour. Si pendant cette période vous mettez très fréquemment à jour votre site Googlebot saura que c'est un site qui est vivant et qui bouge, et qu'on a besoin de lui pour indexer les contenus. Cela est très positif pour vous. À l'inverse, si Googlebot voit que rien ne bouge dans les premiers jours, il aura tendance à plutôt se dire qu'on n'a pas vraiment besoin de lui ici, et il passera moins souvent, voir très rarement. Cela est vraiment négatif. C'est pourquoi pour les premiers jours de votre site, vous devez avoir à cœur de montrer au Googlebot que votre site est réactif, sinon beaucoup de vos efforts pourraient ne pas être vus par Google. Retenez bien ce conseil !

Le référencement se construit sur la durée

Rien n'est instantané

Dès le prochain chapitre, nous commencerons à parler sérieusement de référencement, et je vous enseignerai les rudiments de cette discipline. Dans le futur, vous commencerez à modifier certaines parties de vos sites pour qu'elles plaisent plus à Googlebot, mais sachez que Google ne prendra jamais en considération ces modifications immédiatement.

Le référencement se construit sur la durée, et il s'agit d'un processus long. Il faudra donc attendre que Googlebot repasse sur votre site et prenne en considération les changements pour que Google mette à jour ses résultats de recherche, mais il sera alors fort peu probable que cette mise à jour améliore la position de votre site. En effet, il faut du temps pour construire votre référencement et vous ne grappillerez jamais les premières places immédiatement. Prenez l'exemple du Site du Zéro, c'est un référencement construit sur plus d'une dizaine d'années... alors si vous planchez sur un site qui va donner des cours d'informatique, vous allez vraiment avoir du mal pour déloger le Site du Zéro des premières places !

Les sanctions de Google

Googlebot et les pénalités

Le travail de Google est « simple » : organiser l'information sur le web et la rendre utilisable pour tous. C'est à ce titre que Googlebot va évaluer la qualité de votre site et vous donner des bons points ou des mauvais points. Si votre site respecte les consignes de Google et que vous ne faites que proposer du contenu pertinent et de qualité, Googlebot ne fera que vous donner de bons points et améliorera le positionnement de votre site.

En revanche, si vous ne faites que plagier le contenu des autres sites ou que vous ne respectez pas les consignes de Google, vous risquez d'encaisser des pénalités de la part de Google. Ces pénalités sont très négatives pour votre référencement, plus Google vous donne de pénalités, plus la fréquence de passage du Googlebot diminuera. À terme, Google peut carrément retirer votre site de son index, veillez donc bien à ne pas faire n'importe quoi. 😞

En résumé

- Les robots sont des programmes automatisés arpentant le web tous les jours et remplissant une tâche précise.
- L'un d'eux, Googlebot est un robot d'indexation. Son rôle est de faire parvenir les pages web des sites qu'il visite à Google pour qu'il puisse les ajouter à son index.
- Un moteur de recherche est donc un énorme programme complètement automatisé : aucun humain n'intervient dans le processus d'indexation d'un site.

Partie 2 : Optimisez vos pages web pour le référencement

Cette première partie du cours intitulée « Les actions en interne » traitera de toutes les actions que vous allez pouvoir réaliser sur **votre propre site** pour « l'optimiser niveau référencement ». Elle vous apprendra les bases du référencement, à communiquer avec les robots d'indexation, à penser vos pages d'un point de vue axé "SEO"... Lorsque l'on écrit pour un support quelconque, nous devons forcément obéir à des codes. Pour le web c'est pareil ! Rédiger des pages c'est super, mais encore faut-il rédiger une page valable d'un point de vue référencement pour Google !

Concepts de base d'un site « SEO-Friendly »

Construire un site web c'est passionnant car cela passe par plein d'étapes différentes ! L'une d'elle est la conceptualisation ; on se demande quelles technologies l'on va utiliser sur son site et pourquoi. C'est donc ici que nous allons réellement parler de votre site web pour la première fois, et nous agissons sur sa conception. Pour rappel, dans toute cette seconde partie du tutoriel, nous ne parlerons que de la conception et de l'optimisation de vos propres pages web.

Lorsque vous développez votre site, ne le faites jamais à la vue de Google, nous savons depuis le chapitre précédent que le réflexe de Googlebot sera de traiter votre contenu de manière à l'indexer... la version en développement de votre site sera donc accessible sur Google, et elle sera bien sûr positionnée : ce qui est très négatif pour les raisons que vous imaginez. Agissez toujours en amont de votre projet, et ne déployez que lorsque votre site est tout beau, car il est encore plus difficile après de redresser la barre. 😊

Un site web, du code HTML

A l'exception des sites en Flash dont nous parlerons plus tard, un site web n'est vraiment rien d'autre qu'un ensemble de pages HTML qui sont liées entre elles grâce à la magie des liens. HTML va donc avoir une importance capitale dans votre référencement.



Nous allons nous focaliser ici sur le contenu de la balise `<body>`. Pour le contenu de la balise `<head>` il faudra attendre encore un peu, car c'est plus complexe.

Le langage HTML

Baliser, c'est aider le Googlebot

Le HTML ressemble beaucoup au XML, et pour cause ! Son rôle est simplement de vous aider à créer du contenu web cohérent. Nous savons que le code HTML sera lu par Googlebot et qu'il aura donc une importance capitale, il faudra donc produire un beau code. Lorsque vous avez appris HTML, on vous a forcément fait un topo sur l'histoire de la validité du code des pages web sur le W3C validator, et pour cause : un code valide est un code qui se référence bien. Produire un beau code source garantit une lecture optimale pour le Googlebot, qui sera plus à même de repérer les informations importantes de vos pages web. Commencez donc par créer un site qui soit propre dans son code, et ne vous lancez pas dans la création de sites web si vous n'avez pas **de bases solides en HTML et CSS**.

Les balises HTML ne signifient pas toutes la même chose

Chaque balise a sa signification, et chacune d'elle permet de comprendre la page web. Une balise peut représenter un titre, une autre une image. De la même manière que les balises ne signifient pas toutes la même chose, les informations qu'elles contiennent ne sont pas les mêmes, et elles n'auront pas le même impact sur la manière dont Googlebot va indexer votre contenu. Par exemple, il comprendra que ce qu'il y a dans `` représentera un texte plus important qu'un texte dans une balise ``.

Ainsi, Google sera plus à même de faire apparaître votre site web plus haut pour la recherche « tarte aux pommes » si cette expression se trouve dans une balise `` plutôt que dans une balise `` qui elle, n'a pas vraiment de signification, mais ça, c'est vraiment dans l'absolu. 😊

Le balises de titre

Ces balises sont certainement les plus importantes de toutes, et vous devrez travailler leur contenu le plus possible. Vous les connaissez, il s'agit des balises `<h1>` à `<h6>`. Dans votre stratégie de référencement, vous ne devrez jamais négliger l'importance de ces balises.

Votre `<h1>` est votre **titre de premier niveau**, c'est donc le titre le plus important de la section `<body>`. C'est le titre que le visiteur verra en premier sur votre page web, et il est censé représenter à lui seul l'intérêt de votre page. La dernière norme de HTML baptisée HTML5 introduit de nouvelles balises permettant de structurer vos contenus. La balise `<section>` en fait partie, et comme son nom l'indique, elle permet de définir plusieurs sections dans une page web. La balise de titre `<h1>` est en

effet destinée à être le titre d'une section, vous pouvez donc avoir plusieurs `<h1>` par page web, pourvu que ces titres fassent partie de sections. Il n'y a donc aucun problème à avoir plusieurs `<h1>` page web, ce n'est vraiment pas un drame en soi, mais **cela doit être justifié**.

Vous pouvez donc avoir aussi bien 50 `<section>` avec `<h1>` qu'une seule balise `<section>` avec un seul `<h1>` dans vos pages web. Tant que cela a du sens au niveau du contexte de votre page web et de la sémantique HTML5, Google ne vous en tient pas rigueur. Organisez correctement vos pages web !



N'oubliez jamais d'utiliser une balise `<h1>` dans vos pages web. Une page sans `<h1>` est une page sans avenir.

Si vous ne savez pas ce qu'est une balise `<section>` ou comment l'employer, vous pouvez toujours relire [le tutoriel de HTML5 du Site du Zéro](#). Les sites web se font aujourd'hui en HTML5, vous êtes donc vivement invité à vous mettre à niveau. 😊

Les cas `` et ``

Les balises `` et `` sont également d'une importance capitale dans votre code HTML simplement parce qu'elles permettent de mettre en exergue des expressions que vous jugez importantes, et qui constituent par extension l'intérêt de votre page web. Utilisez ces balises avec modération, et ne mettez pas en valeur 95% de vos textes avec celles-ci. Googlebot est très perfectionné, s'il voit une balise `` il va considérer l'information qu'elle contient comme importante, ne faites donc pas n'importe quoi avec ces balises.



Ces balises servent **UNIQUEMENT** à dire que le texte qu'elles encadrent est important, mais en aucun cas elles signifient « mettre en gras ou en italique ». Ceci se gère en CSS mais n'oubliez pas que Googlebot se moque éperdument de la mise en forme de votre page : il ne la voit pas.

Les textes des liens (Anchor text)

Nous verrons dans la suite de ce cours que les liens vont constituer un enjeu très important dans votre stratégie de référencement. Cependant, nous pouvons déjà commencer à en parler car il est bon de produire des liens corrects sur vos pages web. En effet, les liens se construisent en HTML grâce à la balise `text`.

Dans l'attribut `href=""` vous devrez renseigner la page web visée avant de pouvoir donner un texte au lien juste avant la fermeture de la balise `<a>`. Googlebot, lui, explore les liens, aussi, le texte de votre lien constitue pour lui l'information la plus accessible afin de savoir ce qui se trouve de l'autre côté du lien. Détaillez donc bien le contenu de la page du lien dans le texte afin d'aider Googlebot à mieux s'y retrouver : vous avez des points à y gagner !

L'attribut alt des images

Si vous jetez un œil au code ci-dessus, vous verrez que j'ai renseigné un attribut `alt=""` aux images. Ne pas le renseigner entraîne des pénalités pour votre référencement. Il s'agit là d'une règle fondamentale d'accessibilité, à chaque fois que vous utiliserez une balise `` celle-ci devra comporter un attribut alt. C'est une erreur que l'on voit souvent, **mais qui peut-être très facilement évitée**.

Ratio Contenu/HTML

Pour finir avec ces règles fondamentales de SEO avec HTML, sachez que vos pages devront respecter un certain ratio Contenu/HTML afin d'avoir le plus de chances de plaire à Google. Ainsi lorsque Google va venir analyser vos pages, il faudra qu'au moins 15% de la source de la page soit du contenu pur et pas des balises HTML. Le contenu est ce qu'il y a de plus important aux yeux de Google. Considérez donc cette limite comme la limite minimale : plus le ratio est élevé côté contenu, mieux c'est.



Pour structurer vos pages et rendre votre code HTML plus cohérent, utilisez les balises structurantes fournies par HTML5. Nous avons déjà parlé de `<section>` tout à l'heure, il en existe d'autres comme `<nav>` ou `<footer>`. À l'instar des navigateurs, les moteurs de recherche n'attachent pour le moment que peu d'importance à ces balises, mais il y a fort à parier que la tendance s'inverse dans l'avenir avec l'avènement de HTML5. Prenez donc l'habitude dès maintenant d'utiliser ces balises structurantes qui servent à découper votre code source en espaces logiques, cela aura de fortes chances de plaire au moteur de recherche dans le futur.

Propreté générale du code HTML

Tentez de produire des codes sources qui passeront la validation du W3C, ou au moins qui soient potables, sinon le web ne ressemblera vraiment plus à rien. Travaillez bien vos codes sources et faites le nécessaire afin d'arriver à un bon équilibre entre la page vue par le visiteur et la masse de code source analysée par Google. Plus votre page sera complexe, plus votre code source le sera aussi, et plus Google devra batailler avec vos balises pour indexer votre contenu, l'organe principal de votre site.



Vous n'imaginez pas toutes les horreurs que l'on peut trouver sur le web ; entre les balises qui sont un mélange de BBCode, d'HTML5 et de (x)HTML, sans compter les pages qui se finissent par `</head>`... par pitié, ne vous lancez pas dans la création de sites sur le web si vous ne connaissez pas ses standards. Mettez-vous à niveau avant de continuer ce cours si vous ne vous sentez pas d'attaque. Le [tutoriel HTML5/CSS3 du Site du Zéro](#) est parfaitement adapté aux débutants.



Pour le coup de la page qui se finit par `</head>` oui ! Je peux jurer sur l'honneur que je l'ai vu ! 😊

Une question d'ergonomie

J'espère que vous êtes maintenant sensibilisés à l'importance de votre code HTML. Parlons maintenant d'ergonomie. Il va de soi que votre site doit être ergonomique : s'il est un fait qui a peu changé au cours de l'histoire du web, c'est la nécessité d'une mise en page claire et lisible. On aura beau dire tout ce que l'on veut, les sites web actuels et depuis déjà un bon moment, sont formatés sur un modèle identique.

Ergonomie de base d'un site

Un modèle utilisé partout sur le web

Un site s'articule la majorité du temps sur une structure à 3 niveaux. Dans un premier temps, on trouvera le header avec éventuellement un logo fixé au sommet de toutes les pages *via* une feuille de style. Au bas de ce header, on retrouvera généralement le centre de la page, lui-même divisé en deux parties qui sont le menu et le contenu. Finalement, au bas de la page, on retrouve un footer. Celui-ci ayant pour unique tâche de boucler la boucle, il est généralement chargé des mentions légales du site et autres liens de contact.

Un code source HTML de base

Voici ce que pourrait être le code source HTML5 de base d'une page web moderne. Au passage, vous êtes **vivement** encouragé à toujours utiliser les dernières versions des standards web.

Code : HTML

```
<!DOCTYPE html>
<!-- La section <head>. Nous aurons l'occasion d'en parler plus
tard dans le cours ! -->
<head>
  <title>Le titre de ma page</title>
  <meta http-equiv="Content-Type" content="text/html; charset=iso-
8859-1" />
  <link rel="stylesheet" type="text/css" href="style.css" />
  <link rel="shortcut icon" href="images/favicon.ico" />
</head>

<!-- Mise en page classique avec le header, le corps, et le footer
! Les balises structurantes de
HTML5 sont employées. -->
<body>
  <div id="site">
    <header>L'entête de mon site ! Et bientôt, il y aura une
bannière ou un logo ici !</header>
    <nav id="menu">
      <ul>
        <li>Élément de menu 1</li>
        <li>Élément de menu 2</li>
        <li>Élément de menu 3</li>
        <li>Élément de menu 4</li>
        <li>Élément de menu 5</li>
        <li>Élément de menu 6</li>
        <li>Élément de menu 7</li>
      </ul>
    </nav>
  </div>
</body>
```

```

        </ul>
    </nav>
    <section>
        <article>
            <h1>Le titre de mon article ;) </h1>
            <p>
                Salut, voici le contenu de ma page web 
            </p>
        </article>
    </section>
    <footer>
        <p>
            Le contenu de mon pied de page, mention légale,
etc.
        </p>
    </footer>
</div>
</body>
</html>

```

Ce modèle de base, c'est celui que l'on a tendance à enseigner partout. Ainsi, le cours de HTML5 du Site du Zéro **vous enseignera également cette mise en page de base**, de la même manière que le site [W3Schools](#) ; une sorte de Site du Zéro anglais.



Les quelques innovations qui ont été apportées à ce modèle de base au fil des ans se comptent sur les doigts d'une main. La plus connue est probablement celle de l'ajout d'un menu horizontal à proximité du header du site ; uniquement permise par les avancées technologiques en matière de création d'écrans : ils ne cessent de s'agrandir et de proposer de meilleures résolutions. Autant vous dire que les internautes sont plutôt habitués à cette mise en page sobre mais terriblement efficace.

Être en accord avec la mission de Google !

La mission de Google est d'organiser et de rendre accessible au plus grand nombre l'information sur internet, et plus particulièrement le web. Les internautes étant habitués depuis plus de 20 ans à ce type de mise en page, **c'est tout naturellement de cette manière que Google indexera le mieux votre contenu**. Ainsi, vous avez tout intérêt à opter pour ce type de mise en page, qui se veut ergonomique et qui représente vraiment quelque chose dans l'esprit d'un visiteur. 😊



Étant donné que les sites sont moulés selon ce modèle, on a très vite ressenti le besoin de créer des frameworks CSS pour déployer rapidement une mise en page ergonomique. L'un de ces frameworks est 960 Grid System, qui est très connu et très apprécié. Un autre framework CSS désormais incontournable est le Bootstrap Twitter, un produit de qualité très apprécié des développeurs web.

Question d'ergonomie : le temps passé sur le site

Google est capable de déterminer combien de temps vos visiteurs passent sur votre site. Vous aurez donc à cœur de faire en sorte que vos visiteurs passent le plus de temps possible sur vos pages web, car cela vous rapportera des points pour le SEO. Plus vos visiteurs resteront sur vos pages, mieux elles seront référencées sur Google. Opter pour une mise en page ergonomique, c'est aussi un moyen d'être sûr que le visiteur ne fuira pas dès l'arrivée sur votre site.

La règle d'or des 600px : ergonomie et référencement.

Une règle d'or

À moins d'accéder à votre site via une ancre dans l'URL, le visiteur démarrera toujours sa navigation au sommet d'une page : à hauteur de son header. Le reste de la page étant généralement accessible grâce à un scroll. C'est pourquoi, d'un point de vue ergonomique : **tout doit être dit dans les 600 premiers pixels de votre page web**. J'explique tout de suite ce concept.

Comparons votre site web à un journal

Je vais comparer votre site web à un journal replié, que vous pouvez par exemple lire chez vous en buvant votre café le matin. Comme on peut le voir, ce journal est replié sur lui-même. On dira



ce qu'on voudra : un journal c'est assez gros et pas forcément très évident à tenir en main et à lire. Alors les marchands de journaux se sont rendu compte que pour pouvoir attirer le regard des clients, il fallait plier le journal de telle sorte à ce qu'il paraisse petit et facile à prendre en main. Cela permet en plus un gain de place conséquent sur les étagères.



Les journaux sont pliés de manière à ce que la une soit toujours visible pour le client. Cela va lui donner forcément beaucoup plus envie de lire le journal et donc de l'acheter, puisque généralement dans la une, on a un titre et une introduction avec une belle image bien représentative... Bref ! En voyant la une, on sait directement de quoi l'article va parler, et on va donc **avoir envie de le lire jusqu'au bout** sans pour autant voir l'intégralité de l'article. Comme quoi, une bonne accroche, c'est très important.

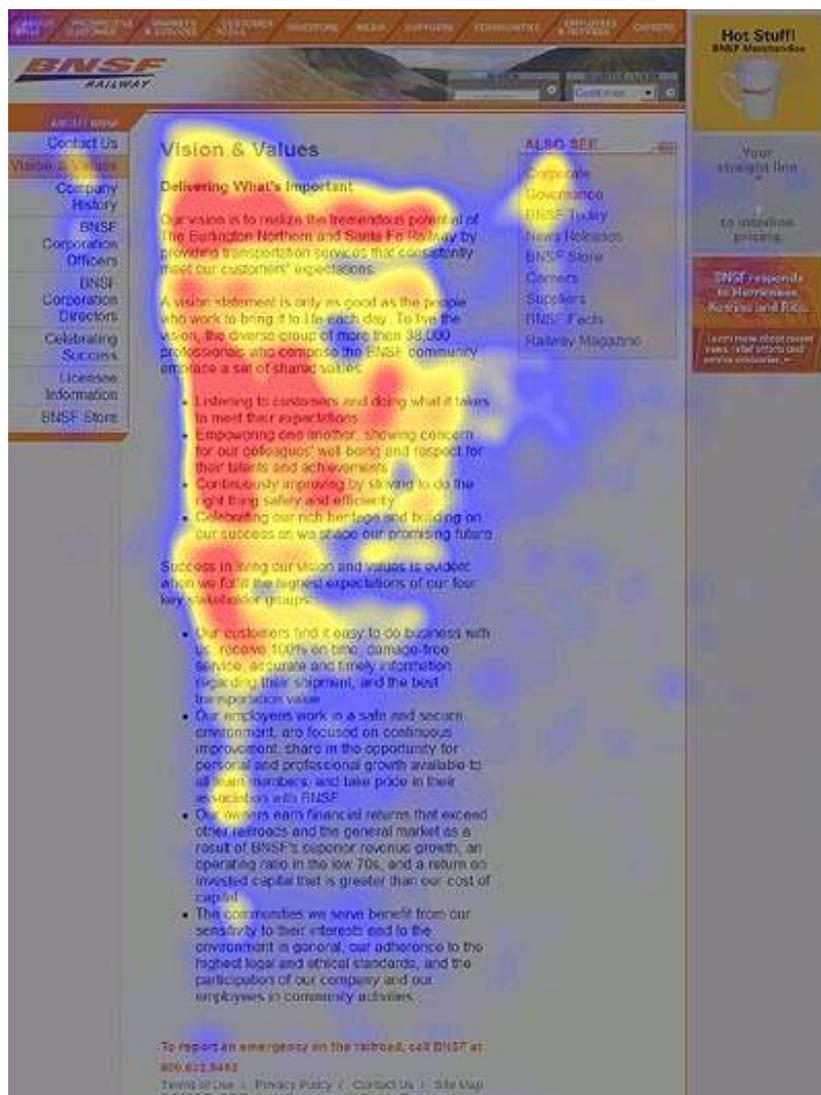
Et bien sur le web c'est un peu pareil. Votre une, ce sont les 600 premiers pixels de chacun de vos articles. À la manière de la une d'un journal : vous **DEVEZ** absolument travailler correctement le début de votre article, pour qu'il soit accrocheur. D'un point de vue typographique et ergonomique, le visiteur aime avoir un gros titre, une introduction de quelques lignes décrivant globalement la problématique de l'article, et ensuite : on se lance. Une fois qu'on arrive au bout de cette accroche, on déplie le journal, et bien sur le web : on scroll. 😊



Voici un test **très intéressant** que vous pouvez faire sur votre site : Google a mis au point un outil nommé [Browsersize](#) qui va vous permettre de vérifier sur votre site quel pourcentage d'utilisateurs va pouvoir visualiser quelle zone de votre site sans avoir besoin de scroller. Vous comprendrez vite pourquoi il faut bien organiser vos 600 premiers pixels ; jouez sur l'opacité dans la zone située dans l'en-tête de l'outil pour révéler les petits secrets de votre mise en page.

Un autre Eye Tracking

Voyez également cette autre étude réalisée cette fois-ci par [Useit](#), qui ressemble beaucoup au Triangle d'Or de Google. Il s'agit du même concept de « carte » : un Eye Tracking, porté cette fois-ci sur un article sur le web.



Vous pouvez voir facilement où va se poser le regard de votre visiteur en regardant cette image : au sommet de celle-ci sur les 600 premiers pixels de l'article. J'imagine que vous avez compris où je voulais venir non ? Travaillez vos 600 premiers pixels. C'est vraiment très important !

La philosophie d'indexation de Google à travers les 600px

Au-delà d'une simple question d'ergonomie ou de confort pour votre visiteur, c'est exactement la philosophie de référencement de Google vis-à-vis de vos visiteurs. Pourquoi ? Parce que Google, lors de l'indexation, va considérer votre site comme une sorte de journal et va se pencher sur la une de votre article : les 600 premiers px. Elle est très importante d'un point de vue SEO étant donné que pour Google, **un mot qui se trouve au début de l'article est bien plus important qu'un mot se trouvant à la fin de cet article**. Le début de l'article correspond bien sûr aux 600px.



Eh oui ! Google va vraiment très loin dans la recherche de votre contenu, et va même "parser" votre code HTML afin de déterminer ce qu'il se passe dans le code source et à quel niveau !

Un autre conseil que je puis vous donner. Il n'est pas rare de voir sur le web des sites dont le footer est occupé par un système de tags. Pour ce genre de système, je dis non, non et **non**. Il n'est rien de plus inutile que de charger son pied de page de mots-clés, si l'on considère la règle d'or que je vous ai donnée plus haut. Les mots se trouvant au bas de la page n'ont presque aucune valeur comparé aux mots se trouvant en haut de la page. Les systèmes de tags sont donc complètement dénués d'intérêts dans les footers. Préférez les systèmes de footer classiques pour votre site. Quelque chose de joli et de bien géré qui finira bien votre page visuellement, en proposant quelques ouvertures avec des liens sur d'autres contenus que votre site propose. Dites vous que n'avez pas à réinventer la roue. 😊



N'oubliez vraiment pas que plus une expression est placée « tôt » dans un article, plus elle aura d'importance. Chargez vos 600 premiers pixels d'expressions intéressantes, la fin de l'article n'a que peu de potentiel par rapport au début !

Pensez à votre visiteur !

Trop de Webmasters ne le font pas, et pourtant... la plupart du temps, c'est quand même à son visiteur que l'on veut plaire non ? Nous allons ici tenter d'avoir un peu d'empathie et de nous glisser dans la peau du visiteur de notre site web et voir ce qu'il attend. Google sait que votre contenu est avant tout destiné aux visiteurs, et il fonde ses algorithmes d'indexation sur cette idée !

Performances, performances, performances !

Un site qui plaît est un site qui va vite

Curieusement, une idée fausse a souvent été de croire que pour plaire à son visiteur, l'essentiel serait d'être le plus original possible dans tous les aspects visuels du site. Pourquoi cela ? Un début de réponse pourrait être que parce que le design est la première chose que le visiteur verra, alors il doit être le plus original possible... ce qui est en très grande partie faux, dans le sens où **originalité rime souvent avec complexité**, lourdeur, et lag à l'affichage.

Pour se démarquer des autres par son design, on aura forcément tendance à rechercher dans le plus complexe : carrousels, animations Flash et une centaine d'images reliées à la CSS. Et ça, c'est une véritable catastrophe pour le visiteur qui lui ne recherche pas originalité, mais confort... Et le confort d'un visiteur se trouve dans l'optimisation des temps de chargement de votre site. Un site qui répond vite est un site qui attire. Lorsque vous demandez quelque chose, vous aimez généralement que la chose vous arrive vite entre les mains non ? Pour le visiteur c'est pareil, lorsqu'il va cliquer sur votre lien Google, il veut vite obtenir ce pour quoi il est venu : l'information, et pas le design.

Ceci a vraiment une importance dans la manière dont Google va indexer vos contenus. Je vous propose [cet article \(EN\) paru en 2010 sur le blog officiel de Google](#) dédié aux Webmasters. Vous y verrez que Google apprécie davantage votre site **si celui-ci est rapide pour le visiteur**.



Google met à disposition des développeurs l'outil [Google Pagespeed](#). Vous pourrez avoir un retour très complet des performances de votre site web sur ce service.

Accélérez vos temps de chargement !

Après cette petite sensibilisation vous vous demandez très probablement comment on optimise les temps de chargement de son site. Je me propose de vous donner quelques exemples ici. Si je ne puis détailler de fond en comble toutes les façons d'optimiser les temps de chargement de vos pages web car il faudrait passer en revue toutes les technologies, je vais vous indiquer

rapidement en quelques points les démarches, en vous conseillant quelques documentations. C'est le moins que je puisse faire.



Utilisez des sprites CSS

Pour chaque image qui compose votre site web, **c'est une requête HTTP en plus qui doit être gérée par le serveur** pour l'afficher au visiteur. Votre article comportera peut-être 10 images, à cela, ajoutons 10 images supplémentaires appartenant au design... déjà 20 requêtes HTTP pour un seul article. Multiplions maintenant tout cela par le nombre de pages vues par chaque visiteur de votre site... cela devient vite affolant. 😊

L'idéal au niveau des CSS, pour améliorer les temps de chargement, est donc de ne pas utiliser d'images du tout pour composer votre design. Les designs « full CSS » s'afficheront extrêmement vite et auront une place toute particulière dans le cœur de votre visiteur et dans celui de Google. C'est quelque chose qui est tout à fait possible et qui commence à se faire de plus en plus sur le web, mais il faut avoir un certain niveau en CSS et être un vrai créatif pour donner quelque chose de bon. **L'alternative à cela est donc d'utiliser des sprites CSS.**



Les sprites CSS de plusieurs grands sites, dont Google.

Les sprites CSS, **c'est l'art de compiler toutes les images qui composent votre site web en une seule.** Ensuite, dans votre CSS, vous n'aurez qu'à afficher cette image et la positionner pour que l'élément graphique que vous souhaitez s'affiche bien à hauteur de la <div> sur laquelle vous travaillez. Ainsi, ce n'est plus 100 requêtes HTTP qui vont servir à récupérer toutes vos images, mais une seule. Des économies énormes pour votre serveur.



Il existe **un tutoriel très bien réalisé** sur le Site du Zéro pour vous aider à appréhender les sprites CSS. Vous verrez que ce n'est franchement pas complexe à mettre en œuvre.

Pour information, le Bootstrap Twitter a été construit grâce aux sprites CSS. En effet, il est composé de classes CSS3 très développées qui donnent un excellent rendu, et les quelques petites icônes **que le framework propose sont gérées par des sprites.** Il peut constituer en cela une bonne base pour commencer un site.

Placez convenablement vos codes JavaScript !

JavaScript est un langage généralement utilisé dans des pages web. Malgré des idées reçues, JavaScript est un langage vraiment puissant qui vaut le détour. Il s'exécute côté client, **ce qui signifie que votre serveur n'aura pas à l'interpréter** : déjà un bon point : c'est au moteur JavaScript de votre navigateur de le gérer.

JavaScript est puissant, mais il est généralement mal utilisé. Vous pouvez soit choisir d'en mettre directement dans votre page web, qui est bien souvent une mauvaise solution pour la lisibilité de votre code source par Google et vous-même, ou soit le placer dans un fichier qui portera l'extension .js **que vous prendrez soin de rattacher avec votre page web grâce à la balise script :**

```
<script src="fichier.js"></script>
```

Vous devez inclure un JavaScript soit au tout début de votre page, dans la balise `<head>` soit à la toute fin, juste avant que l'on ferme la balise `<body>` avec `</body>`, cette dernière solution étant de loin la meilleure. Ceci est très important car du JavaScript chargé en plein milieu d'une page web alourdira inutilement l'ensemble, il est préférable de charger du JavaScript juste avant la fermeture de la balise `<body>`, ainsi le contenu du site sera entièrement chargé quand JavaScript entrera en jeu. Je puis vous rediriger par ailleurs sur le [cours de JavaScript](#) du Site du Zéro qui devrait vous aider à bien comprendre cette technologie.

Vos requêtes SQL

Le cours de [PHP/MySQL](#) sur le Site du Zéro est introduit par ces lignes : "*Blogs, réseaux sociaux, pages d'accueil personnalisables... Depuis quelques années, les sites web ont gagné en fonctionnalités et sont devenus dans le même temps de plus en plus complexes.*"

Que c'est vrai... Votre site n'est plus le simple code source HTML et les 50 lignes de CSS qu'il pouvait être autrefois. Aujourd'hui, c'est une véritable machine intelligente qui fonctionne silencieusement tous les jours de l'année. Quel que soit le langage employé ; c'est tout un univers de variables et de fonctions qui s'articule... **mais aussi de requêtes SQL.**



On commence à plutôt bien se connaître vous et moi maintenant, aussi je pense que vous avez commencé à remarquer que j'adore les statistiques et les chiffres. Savez-vous qu'un site web utilisant une base de données est en moyenne 5 fois plus lent qu'un site uniquement construit sur des fichiers ? Ça paraît fou, et pourtant... Si l'on peut difficilement se passer de base de données, tant elles sont importantes pour avoir un site un minimum intéressant, on peut toujours réduire les temps de chargements en diminuant le nombre de requêtes SQL, en les optimisant, **et en fournissant des Index aux tables** .

Couplez tous ces conseils avec toutes les indications sur l'ergonomie que je vous ai données précédemment, et cela devrait donner quelque chose de convenable en terme de temps de chargement.



Un utilitaire nommé [Minify](#) codé en PHP commence à se faire une bonne réputation dans la communauté ; son principe est de "nettoyer" automatiquement vos codes sources en plaçant correctement vos JavaScript ou en fusionnant vos CSS en une seule. Pour l'avoir essayé, je le trouve plutôt bon. C'est peut-être susceptible de vous intéresser. En revanche, ne l'utilisez vraiment que lorsque vous savez précisément ce que vous faites... son utilisation pourrait avoir des effets pervers et se retourner contre vous en ayant exactement l'effet inverse sur votre site.

Google et l'empathie

Pensez à vos visiteurs, pas aux robots

Google donne pour conseil aux Webmasters de ne pas chercher à travailler le référencement mais de toujours proposer un contenu et un site qui plaisent aux visiteurs, avant de chercher à plaire au robot. Gardez toujours cette idée en tête, mais n'oubliez jamais d'intégrer une dimension à SEO à vos développements (au sens large) de sites web.

Je ne cherche par ailleurs à vous imposer aucune mise en page ou contrainte : c'est votre site et vous en faites ce que vous voulez. Retenez simplement qu'un carrousel en header est plus beau qu'un logo, mais qu'il demande aussi plus de code HTML ; code que Googlebot considère comme relativement peu intéressant à côté de votre vrai contenu, et cela va lui compliquer la vie plus qu'autre chose, en plus de faire baisser votre ratio code/contenu. 😊

En résumé

- Un site web se conçoit en HTML avant tout, et de préférence en HTML5. Vous avez intérêt à avoir le code le plus propre possible.
- Ne négligez jamais vos balises de titre et jouez sur l'importance de vos expressions pour faire ressortir l'intérêt de votre contenu.
- Un mot en haut de page est plus important qu'un mot en bas de page.
- Dans la mesure du possible, optez pour un site web fondé sur le modèle ergonomique de base !
- Un site qui se charge vite est un site qui plaît.

Communiquez avec Googlebot !

Googlebot ne dort jamais. Il travaille tous les jours à récupérer des informations sur votre site pour permettre son indexation. S'il est vrai que c'est un véritable acharné du travail, il est aussi vrai qu'il aime qu'on lui donne des indications sur ce qu'il doit faire sur votre site, par exemple lui dire quelle page indexer, et quelle page ne pas indexer. Ou encore quel lien suivre, et quel lien ne pas suivre... Vous n'aimeriez pas par exemple que vos pages d'administration soient visibles sur Google.

Pour communiquer avec Googlebot et lui donner ces infos, il va falloir passer par le contenu de la section **<head>** de votre page web. Il n'y aura pas 36 solutions, nous allons devoir passer par les balises **<meta />**. Oui ! Ces balises qui étaient totalement indispensables pour référencer son site dans les années 1990 sont toujours là !

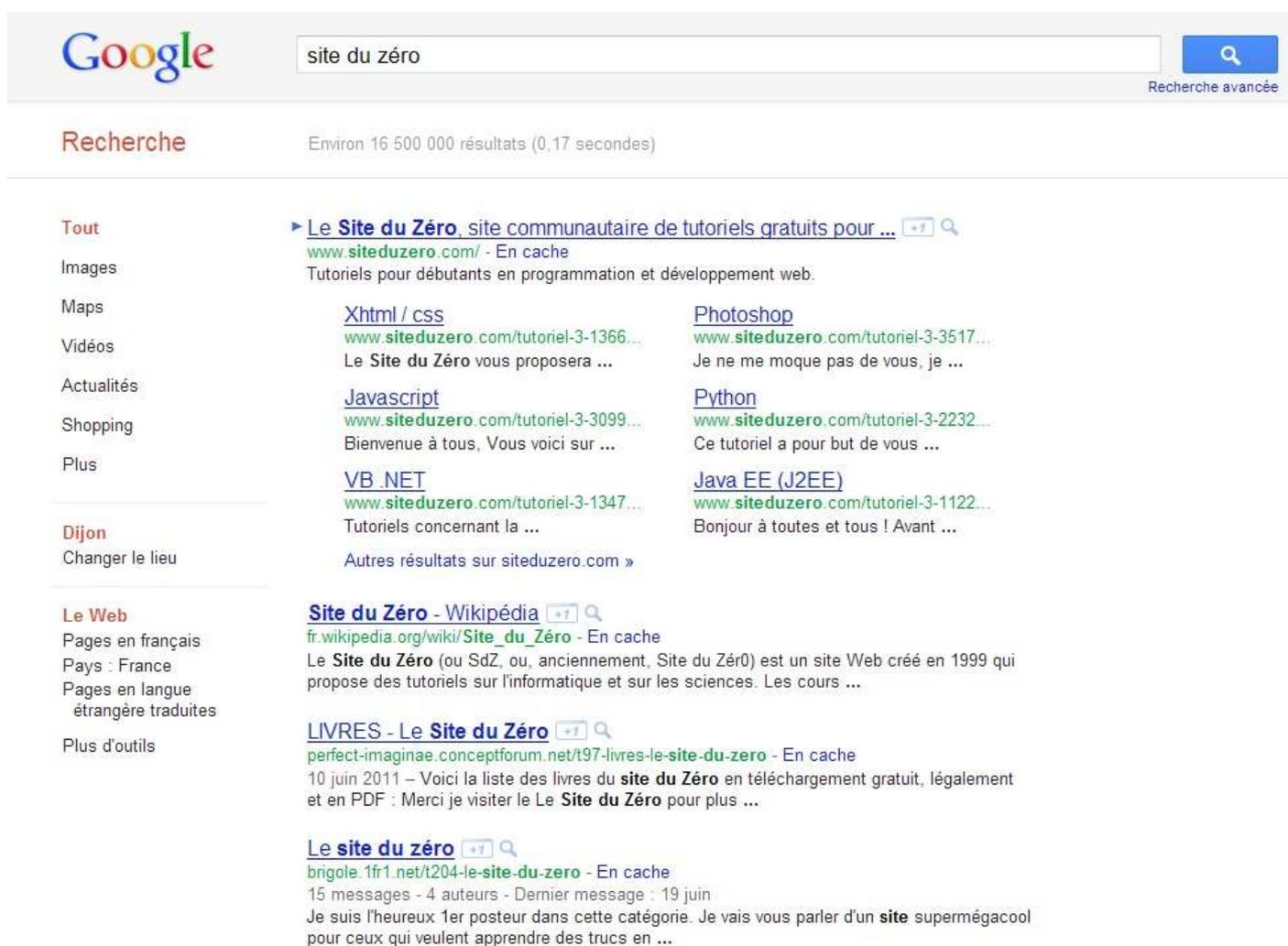
Aidez Googlebot à générer le résultat de recherche

Il existe des tonnes d'éléments intéressants dans votre code source HTML. L'un d'eux est un peu à part, et doit être exploité par Google : c'est la balise **<title>**. Elle doit être totalement unique au sein de votre page web.

Google génère ses résultats à partir des pages web

<title> : une balise capitale

La balise **<title>** dispose d'une importance primordiale parce qu'elle va servir à Google à générer le résultat de votre site.



The screenshot shows a Google search interface with the query "site du zéro". The search results are displayed in a grid format. The first result is "Le Site du Zéro, site communautaire de tutoriels gratuits pour ..." from www.siteduzero.com. Below this, there are several sub-results for various topics like Xhtml/css, Photoshop, Javascript, Python, VB.NET, and Java EE (J2EE). The second main result is "Site du Zéro - Wikipédia" from fr.wikipedia.org/wiki/Site_du_Zéro. The third main result is "LIVRES - Le Site du Zéro" from perfect-imaginae.conceptforum.net. The fourth main result is "Le site du zéro" from brigole.1fr1.net. Each result includes a title, URL, and a brief description.

Nous pouvons voir assez facilement que chacun des résultats sur cette page est présenté selon un même *pattern*. En effet à chaque fois, on trouvera **un titre avec un lien qui est fait à une page web**. On aura également l'URL de cette page juste en-dessous dans une police verte, et en troisième position, nous avons une courte description de la page en question. Je pense que je vous ai mis la puce à l'oreille sur l'utilité de cette balise **<title>**... le contenu de cette balise va être réutilisé par Google pour générer ses résultats, et en particulier la première ligne : **le lien**.

Ainsi, si cette balise n'est pas bien travaillée, non seulement Google ne placera pas bien la page dans ses résultats, mais en plus, pour les quelques personnes qui la verront, elle ne sera pas attractive et personne ne cliquera dessus.

Le Site du Zéro pour exemple

Pour continuer à expliquer ce qu'est la balise `<title>` nous allons prendre l'exemple de celle de la page d'accueil du Site du Zéro. Placez-vous donc sur la page d'accueil du Site du Zéro et utilisez Ctrl + U vous y verrez le `<title>`.

Code : HTML

```
<!--Doctype et ouverture du <head> -->
<title>
    Le Site du Zéro, site communautaire de tutoriels gratuits
    pour débutants : programmation, création de sites Web, Linux...
</title>
<!--fin du <head> puis <body>... -->
```

Le contenu de ce qui se trouve dans `<title>` est réutilisé par Google pour former le lien qui mènera vers votre page. Vous pouvez comparer le code source ci-dessus, avec le visuel de la page de résultats pour vous en rendre compte. Vous devez maintenant comprendre que la balise `<title>` est d'une importance vraiment capitale dans votre référencement. Il faudra travailler **vraiment** vos `<title>`.

Choisir un bon titre

Vous **ne pouvez pas** indiquer n'importe quoi comme `<title>` pour votre page web. Celui-ci doit être riche en mots-clés ; mots-clés que l'internaute tapera dans la barre de recherche de Google afin de trouver votre page web. Votre `<title>` doit être en adéquation avec votre article. N'essayez pas d'être original dans votre titre, tentez simplement de désigner votre article par une expression de quelques mots, cela est généralement très bien. Par ailleurs, tentez de ne pas dépasser les 60 caractères. Dans un `<title>` on est direct et précis.



Bien sûr, le choix d'un bon `<title>` est quelque chose de beaucoup plus complexe que cela. Je ne peux pas vous en parler plus en détail pour le moment de peur de vous embrouiller, mais sachez qu'il existe un outil nommé **Google Insights** qui vous donnera des informations essentielles afin d'arriver à choisir les bons mots-clés pour votre `<title>`. Pour le moment on fait simple, mais sachez quand même que par la suite, nous allons nous attaquer à l'utilisation de **Google Insights**, et nous verrons à ce moment-là en profondeur comment choisir un bon `<title>`.

Pour le moment, tentez d'imaginer ce qu'un utilisateur pourrait taper comme mots-clés dans la barre de recherche pour trouver votre page web, et essayez de construire votre `<title>` en conséquence. J'avoue que ce n'est pas très propre, mais pour le moment ça ira.



N'oubliez **JAMAIS** de mettre un `<title>` à vos pages. C'est comme le `<h1>`, il ne faut pas l'oublier ! Par défaut, si votre page ne possède pas de `<title>`, l'URL de celle-ci est généralement considérée comme son `<title>` : c'est très mauvais pour votre référencement.

Les meta description et keywords ne sont pas encore morts !

Dans les années 1990, le moteur de recherche qui dominait les débats, **Altavista**, fondait son succès sur un algorithme d'indexation des pages qui se basait sur les balises `<meta />` et leur contenu dans la zone `<head>` de chaque page Web. Si Google a totalement bouleversé ce système, il s'appuie encore aujourd'hui sur **certaines** balises `<meta />` pour obtenir le plus rapidement possible des informations sur notre site. Si Google a passé sous silence ces balises, il y en a encore deux qui sont très utilisées.

Deux balises `<meta />` qui survivent aux années

Votre contenu, votre centre d'intérêt

Que l'on soit bien clair là-dessus, pour Google votre contenu est vraiment **le point central** de votre site. Ainsi, c'est sur ce plan-là que vous aurez le plus de points à gagner au niveau de votre référencement. Cependant, si Google ne se focalise presque plus du tout sur les balises `<meta />`, il en prend encore quelques-unes en compte. Et par "quelques-unes" j'entends vraiment deux ou trois balises, pas plus. 🐼



N'oubliez pas qu'une métadonnée est par définition une donnée qui définit une donnée. Même si Google a passé sous silence la très grande majorité d'entre elles, on n'a pas encore pu inventer mieux que les quelques métadonnées que l'on va voir dans ce chapitre pour communiquer rapidement des infos sur sa page à Googlebot.

À l'époque d'Altavista, on avait une vingtaine de balises `<meta />` qui servaient à elles seules à bien référencer son site. À l'époque Google, on n'en a plus que deux majeures, à savoir les balises `meta keywords` et `meta description`. Commençons par la balise `<meta name="description" />` qui est la plus intéressante. 😊

META description

Décrivez votre contenu !

La balise `<meta name="description" />` a elle aussi son petit rôle à jouer dans l'indexation de votre page web. Vous l'aviez peut-être remarqué, mais le contenu de cette balise va également s'afficher dans les résultats de recherche sur Google. Juste en dessous de l'URL du site Web, en troisième position.

▶ [Le Site du Zéro, site communautaire de tutoriels gratuits pour ...](#) 🔍
[www.siteduzero.com/](#) - En cache
Tutoriels pour débutants en programmation et développement web.

Cette balise est là pour que vous puissiez décrire plus facilement votre contenu au Googlebot afin que celui-ci indexe mieux votre page web. Comme d'habitude, il faut que cette description soit riche en mots-clés, tout en étant attractive pour le visiteur pour vous démarquer des autres pages dans les résultats. Cette description obéit également à quelques codes. S'il est généralement conseillé de ne pas dépasser les 60 caractères pour son `<title>` ; pour la description, la limite est à 200 caractères. À vous d'être précis maintenant.



Et dis, comment je la place cette description moi ?

Elle se place exactement comme la balise `<title>`, dans la balise `<head>`. Ouvrez simplement une balise `<meta />` avec les attributs `name="description"` et `content="votre description"` comme ceci :

Code : HTML

```
<meta name="description" content="C'est la description de ma page !" />
```

Vous pouvez par exemple placer cette ligne juste en dessous de votre balise `<title>`, elle peut vous faire gagner des points au niveau du référencement, et elle est également là pour jouer un rôle d'accroche pour l'internaute qui regardera les résultats de recherche Google.

META keywords

Les mots-clés de votre page !

Alors elle, **c'est LA balise qui a été la plus grande perdante de ces 10 dernières années**. À l'époque d'Altavista, elle était aussi importante que votre contenu peut l'être aujourd'hui, c'est vous dire. Les gens qui s'occupaient du référencement de leur site se contentaient généralement de bourrer le contenu de cette balise de mots-clés, et hop, cela suffisait à s'assurer une bonne place dans les moteurs de recherche. Aujourd'hui, c'est une balise dont l'importance a été terriblement diminuée, à tel point qu'elle ne sert aujourd'hui pratiquement à... rien.

Ce n'est pas très juste de dire ça, mais il faut bien comprendre que Google ne se focalise pratiquement plus du tout dessus. L'intégrer à vos pages peut être une bonne chose si vous le souhaitez et c'est même recommandé. Mais cela n'influera quasiment pas sur votre référencement. Pour la placer, c'est exactement comme la balise `meta description`.

Code : HTML

```
<meta name="keywords" content="mots-clés" />
```

Dedans, vous avez simplement à indiquer des mots-clés, en les séparant à chaque fois par une virgule. Le Site du Zéro utilise encore une fois la balise `meta keywords`, vous pouvez prendre exemple dessus afin de construire la vôtre. Voici la balise `meta keywords` de la page d'accueil du Site du Zéro.

Code : HTML

```
<meta name="keywords" content="Cours, tutoriel, didacticiel,
informatique, programmation, c, c++, php, ..." />
```

Encore une fois, si vous souhaitez placer de bons keywords dans cette balise `<meta />`, il faudra attendre encore quelques chapitres.

Voilà comment vont s'articuler vos balises `<meta />` ainsi que votre balise `<title>` afin de permettre au Googlebot d'analyser votre site un peu plus facilement, et de votre côté d'assurer quelques points supplémentaires pour votre site au niveau de votre référencement.

Question utilité ?

Ces balises ne servent pas ou peu

- **Meta description :**

La balise meta description est encore relativement intéressante puisque son contenu apparaît dans la page de résultat de Google. Vous devriez y renseigner quelques mots-clés et tenter d'avoir une belle accroche pour votre visiteur. Néanmoins, lorsqu'elle n'est pas utilisée, Google est capable d'identifier d'autres éléments de votre page comme la description.

- **Meta keywords**

: L'époque Altavista est terminée mes amis ! Cette balise ne sert à rien, mais il peut être quand même intéressant de la renseigner. À vous de voir... mais vraiment, on ne vous tiendra jamais rigueur de la non-utilisation de cette balise.

Interdire l'indexation d'une page avec le meta robots !

Nous savons donc maintenant que nous pouvons utiliser les balises citées ci-dessus afin de décrire un peu notre contenu au Googlebot. Cependant, parfois vous aimeriez que Googlebot n'indexe carrément pas certaines pages de votre site... et là c'est difficile, étant donné que vous ne savez pas du tout comment vous y prendre pour lui dire cela.

Un nouveau meta : Robots

Gérer l'indexation d'une page

Le `meta robots` est un meta auquel Googlebot, ainsi qu'un très grand nombre de robots d'indexation (pour ne pas dire tous) sur le web vont se référer afin d'adopter un certain comportement vis-à-vis de votre site. Vous avez la possibilité de dire simplement aux robots, dont Googlebot, de ne pas indexer du tout votre page grâce à ce meta. C'est très simple, le `meta robots` se forme de la manière suivante, à l'intérieur de votre balise `<head>`.



Code : HTML

```
<meta name="robots" content="Instruction1, Instruction2" />
```

Voilà par exemple le type d'instruction que l'on va pouvoir utiliser dans le `meta robots`.

Code : HTML

```
<meta name="robots" content="index, follow" />
```

Index, follow signifie que l'on va pouvoir autoriser Google à **indexer** notre contenu **ET** à **suivre** les liens présents sur la page.



C'est le comportement que Googlebot adopte par défaut si vous ne définissez pas de meta robots.

Ainsi, avec cette configuration, Googlebot va dans un premier temps récupérer les informations sur votre site dans le but de l'indexer et va en plus de cela, **suivre** les liens `` présents sur votre page. Il va alors vérifier dans cette nouvelle page à laquelle il vient d'accéder s'il peut l'indexer ou non. Pour info, *To follow* signifie *Suivre* dans la langue de Shakespeare. Nous ce que l'on veut, ce n'est pas ça. On veut que le Googlebot n'indexe pas notre page, et ne suive pas les liens présents sur celle-ci. Pour cela, on va très simplement lui indiquer l'information `noindex, nofollow`.

Code : HTML

```
<meta name="robots" content="noindex,nofollow" />
```

C'était simple non ? En voyant cela, Googlebot va simplement se dire ceci :

Citation : Pensées d'un Googlebot...

"Ah ! J'ai un noindex dans la balise meta robots. Bon j'indexe pas la page, mais au cas où il y aurait un lien vers une autre page ici, je vais quand même jeter un coup d'oeil.

...

Ah voilà un lien ! Bon, je vais voir ce que je trouve de l'autre côté... ah bah non. Il y aussi un nofollow à côté du noindex dans le meta robots... Bon bah, j'ai rien à faire ici moi du coup, a+ !"

C'est ce qu'il faut retenir.

Vous pouvez également faire d'autres combinaisons comme par exemple mettre une `noindex, follow`. Là Googlebot n'indexera pas votre contenu, mais ira quand même visiter les pages qui sont "linkées" histoire de voir s'il y trouve des contenus à indexer.



Lorsque vous êtes en « index, follow » vous pouvez ajouter également l'instruction "all". C'est une valeur par défaut qui dit au Googlebot de tout suivre et de tout indexer. Elle est interprétée par le Googlebot implicitement lorsque vous ne la marquez pas, à vous de voir si vous voulez la rajouter ou non. Personnellement je le fais. Vous avez également l'instruction « none » qui est l'inverse du « all ». Là Googlebot sait directement qu'il ne doit absolument rien faire avec cette page.

Gérer le suivi des liens au cas par cas

Définir le suivi d'un lien directement dans la balise <a>

Admettons que dans votre page vous ayez plusieurs liens. Vous souhaitez que Googlebot en suive certains, mais pas d'autres. Pour cela c'est simple, on a bien le `follow/nofollow` mais celui-ci se place dans le `<head>` et quelle que soit la valeur que vous choisissez, elle va s'appliquer à tous les liens présents sur la page. L'idéal serait donc d'arriver à placer un `follow/nofollow` sur le lien lui-même, et devinez quoi ? C'est possible.

Cela se fait très facilement grâce à l'attribut `rel=""` :

Code : HTML

```
<a href="une-page.html" rel="follow" />Une page<a> <!-- Googlebot va
```

```
suivre ce lien ! -->
<a href="autre-page.html" rel="nofollow" />Une autre page<a> <!-- Il
ne le suivra pas celui-ci !-->
```

C'est facile non ? Admettons que vous ayez un lien qui mène à vos pages d'administration sur votre site. Vous souhaiteriez que ce lien ne soit pas suivi par Googlebot alors faites simplement :

Code : HTML

```
<a href="admin.html" rel="nofollow">Admin</a>
```

Et dans admin.html, il faudra bien que vous pensiez à marquer la page en noindex (au moins) dans le meta robots.



J'espère quand même que vous avez déjà le réflexe de cacher ce genre de liens avec du PHP ou autre... 🤪

Interdire l'indexation d'un site avec le robots.txt !

Vous savez interdire/autoriser l'accès de certaines pages, mais à moins d'utiliser un système automatique pour afficher le même meta robots sur toutes vos pages vous allez avoir du mal à gérer ça pour plusieurs pages, pour un répertoire ou même tout votre site. Je vous propose une solution à cela, le robots.txt.

Gérez les accès des robots avec robots.txt

Un fichier à créer sur votre site

Les robots, à leur arrivée dans votre site vont regarder si un fichier appelé "robots.txt" est présent à la racine de votre site. Ce fichier va servir à fournir des informations sur votre site pour les robots dans le Googlebot. Le robots.txt est un fichier .txt tout bête que vous pouvez créer sur votre ordinateur très simplement à l'aide du bloc-note.



Ce système est sensible à la casse, le fichier devra s'appeler « robots.txt » en minuscule avec un s, et pas autrement. Il devra également se trouver à la racine de votre site et **pas ailleurs**, sinon les bots ne le trouveront pas et il ne servira à rien.

Voyez par exemple le [robots.txt](#) du Site du Zéro, ou encore [celui de Google](#). Un robots.txt vous servira généralement à interdire ou à autoriser l'indexation de certaines pages de votre site à certains robots ou à tous.

Écrire son robots.txt

Généralement, on commence son robots.txt en indiquant le User-agent qui est visé. Le User-agent est un nom qui va vous permettre d'identifier un robot par rapport à un autre. Par exemple, voici le user-agent de Googlebot.

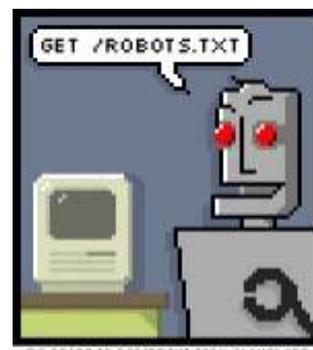
Citation : User-agent de Googlebot

```
Googlebot/2.1 (+http://www.google.com/bot.html)
```

Ainsi, pour commencer à écrire votre robots.txt, vous allez devoir renseigner dans un premier temps le User-agent.

Code : Autre

```
User-agent : Googlebot/2.1 (+http://www.google.com/bot.html)
```



Ainsi, toutes les instructions qui vont suivre vont concerner le Googlebot et pas les autres robots. On peut également décider de

viser tous les robots avec le signe *.

Code : Autre

```
User-agent : *
```

Là, ce sont tous les robots qui seront touchés. Ensuite, vous pouvez utiliser les instructions **Allow** : et **Disallow** : en les faisant suivre du lien du fichier ou répertoire afin d'autoriser ou non son indexation.

Code : Autre

```
User-agent : *  
Allow : /index.php  
Disallow : /admin
```

Dans ce robots.txt j'autorise tous les bots à indexer la page `index.php`, mais j'interdis aussi à tous les robots d'indexer le répertoire `/admin`. On va considérer dans cet exemple que `/admin` contient tout le système d'administration du site. Le conseil que je vous donne est d'utiliser toujours le * pour viser tous les différents robots. En effet, si vous ciblez seulement le Googlebot en lui interdisant d'indexer votre répertoire `/admin`, cette interdiction ne touchera pas les robots de Yahoo! par exemple. Faites bien attention à cela.



Je ne vous recommande pas d'interdire l'indexation de répertoire sensibles : faites attention car le robots.txt est accessible au public, il suffit de taper son URL pour le lire. Si vous interdisez ici l'indexation de répertoires d'administration, vos visiteurs seront capables de le retrouver. Veillez donc bien à utiliser des vérifications et même à interdire l'indexation des pages d'administration directement dans le code HTML de ces pages grâce au meta robots mais pas dans le robots.txt.

On peut de cette manière autoriser ou interdire facilement l'indexation de n'importe quel contenu sur son site. Si aucune instruction n'est définie dans le robots.txt pour une page appelée, par exemple `test.html`, le Googlebot va se référer à ce qui se trouve dans la balise meta robots. Et si cette balise n'existe pas, il l'indexera. Dites-vous toujours que si vous ne spécifiez rien, Google indexe par défaut.



Pensez bien à être cohérent. N'allez pas autoriser l'indexation d'une page dans le robots.txt alors que dans la page en question, vous avez indiqué un `noindex` en meta robots.

Vous pourrez trouver une liste exhaustive des User-agent sur le site web très bien construit User-Agents.org. 😊

Bonne pratique

Renseigner un robots.txt dans tous les cas

Les robots chercheront toujours à obtenir le fichier robots.txt qui se trouvera à la racine de votre site web, une bonne pratique consiste donc à toujours fournir ce fichier même si vous décidez de n'inclure aucune règle d'indexation particulière à l'intérieur. Vous devrez donc créer le fichier robots.txt et y placer le contenu suivant.

Code : Autre

```
User-agent : *  
Disallow :
```

Vous remarquerez que rien n'est placé derrière la directive `Disallow`, et pour cause, on fait simplement le choix ici de ne rien interdire. Les robots vont ensuite gérer les pages web de votre site au cas par cas, selon ce que vous y avez renseigné dans votre balise `<meta name="robots" />` !

En résumé :

- Utilisez la balise `<meta /> robots` pour interdire l'indexation de certains contenus.
- La balise `<title>` et la balise `<meta />description` ne sont à négliger sous aucun prétexte.
- La balise `<meta /> keywords` ne présente plus grand intérêt aujourd'hui.
- Ces données doivent être chargées de mots-clés. Vous apprendrez à les construire correctement plus tard.
- Le fichier `robots.txt` se veut être l'équivalent de la balise `<meta /> robots`, mais à l'échelle du site entier.
- Le fichier `robots.txt` se place à la racine de votre site, et nulle part ailleurs.

Un site web, une arborescence, des URL ! (1/2)

Parlons de l'arborescence de votre site. Qui dit arborescence dit URL. Nous verrons dans ce chapitre que les URL qui mènent vers votre site sont avant toute chose un moyen de localiser vos pages sur le web, mais qu'elles ne constituent pas un critère d'indexation à proprement parler comme peut l'être la balise `<title>`, et ce, depuis la mise à jour Google Panda. Pourquoi en parler dans un tutoriel sur le référencement alors ? Tout simplement car il est recommandé de traiter ses URL proprement : si le robot s'en moque, avoir des URL propres reste une façon de plaire à votre visiteur et de l'inciter à cliquer sur l'URL tout en conservant une organisation claire dans vos contenus !

 Notez qu'avant Google Panda, l'impact des URL dans le référencement ne cessait de faiblir au fil des années. Disons simplement qu'elles ont été mises au placard définitivement depuis cette mise à jour. 😊

La base de l'URL : le nom de domaine ! Les URL de votre sites

Le nom de domaine

Même si Google ne prend pas réellement en compte vos URL, il y distingue bien le nom de domaine, **qui reste le seul élément pris en compte dans l'indexation**. Il est donc important de choisir un bon nom de domaine. Pour rappel, le nom de domaine est en fait la première partie de l'adresse de votre site web. C'est une chaîne de caractères qui se termine généralement par « .com », « .fr » ou « .org ». Pour l'exemple du Site du Zéro : son nom de domaine est **siteduzero.com**. Le « www. » étant ce que l'on appelle un sous-domaine.

Les noms de domaines se composent de deux éléments

siteduzero.com

- La partie **bleue** est ce que l'on appelle le domaine.
- La partie **verte** est ce que l'on appelle l'extension. Pour votre culture, on appelle ces extensions des TLD : des domaines de premier niveau.

Un nom de domaine est donc l'empreinte unique d'un site web. C'est là le premier champ de bataille entre les sites web pour gagner des points **au niveau du référencement** et se démarquer par rapport à son concurrent. Le choix du nom de domaine est quelque chose de primordial puisque le visiteur le verra sur chacune de vos pages dans sa barre d'adresse. Nous allons donc voir ici comment choisir un nom de domaine « SEO-Friendly ». 😊

Le nom de domaine est le nom de votre site

La première chose à faire est de choisir un nom de domaine en adéquation avec son projet de site. En effet, si vous donnez un certain nom à votre site, votre nom de domaine devra également porter ce nom. Cela paraît vraiment logique, mais il m'est arrivé de voir des sites qui ne le faisaient pas. Au-delà de la simple question de confort pour le visiteur cela a **un réel enjeu pour le référencement**. Dans le sens où si nom de domaine et nom du site sont identiques et contiennent les deux un mot-clé intéressant par rapport à votre thématique, **c'est déjà un bon point**.

On a l'habitude de placer un mot-clé en rapport direct avec la thématique du site. Pour reparler un peu des jeux-vidéo Zelda, si vous voulez créer un site qui va donner des solutions sur les principaux volets de la série, des news et tout ce qui va avec, essayez d'inclure le mot « zelda » dans votre nom de domaine parce qu'il a de fortes chances d'être tapé par un utilisateur lors de sa recherche, cela fera déjà un excellent mot-clé pour chacune des pages de votre site. Vous verrez tout l'intérêt de ceci lorsque nous aborderons la longue traîne.

 Faites donc en sorte de sélectionner un nom de domaine qui soit identique au nom de votre site, et si en plus de cela, il peut contenir un mot-clé caractéristique de votre thématique, c'est parfait. 😊

L'extension du nom de domaine

Même si votre extension n'a pas grand intérêt au niveau du SEO, nous pouvons quand même en parler un tout petit peu, car vous devez probablement vous demander quoi choisir : *.com*, *.fr*,



.org ?...Vis à vis de Google, la réponse n'a pas grand intérêt. Matt Cutts [revenait sur cette question \(EN\)](#) il n'y a pas longtemps sur l'un de ses posts Google+ à l'occasion des nouvelles extensions qui sont apparues récemment.



Il y dit que les systèmes d'indexation Google sont suffisamment avancés pour repérer la langue dans laquelle chaque page web est écrite et que Google ne s'arrête donc pas seulement à l'extension du nom de domaine. Partant de ce fait, vous allez pouvoir réserver un nom de domaine en sachant que l'extension ne représente au final pas grand-chose pour le robot. Pour ma part, je vous conseillerai quand même de choisir une extension en fonction de la langue que cible votre site. Si votre site se veut 100% français, selon moi, une extension en.fr « est » toute indiquée. Si en revanche vous voulez proposer un site dans plusieurs langues qui pourra être visité internationalement, vous pouvez vous laisser tenter par un nom de domaine en .com. C'est assez libre de ce côté-là au niveau du référencement, vu que l'on sait que le robot ira identifier la langue utilisée directement au contact de votre contenu.

Notez quand même qu'il existe un réel intérêt à utiliser une extension en « .fr » si votre site est 100% français. En effet, dans ce cas de figure, votre site ressortira de manière prioritaire dans les résultats de recherche de la version française de Google par rapport aux sites en .uk, qui eux, seront prioritaires dans la version anglaise (Royaume-Uni) de Google. Le choix de votre TLD est libre, mais gardez cette petite remarque en tête.

Âge du nom de domaine

L'âge de votre nom de domaine a également un impact sur votre référencement. En effet, plus le nom de domaine est ancien, plus le site web sur lequel il pointe est considéré comme étant de confiance par Google. Il peut être une bonne idée de réserver votre nom de domaine le plus tôt possible, mais ceci n'a qu'un impact très mineur sur votre référencement.

Longueur du nom de domaine, un dernier conseil

Pour clore cette partie sur les noms de domaine, nous allons enfin finir par évoquer la longueur idéale qu'un nom de domaine doit avoir. Eh oui ! Un nom de domaine doit être ni trop court, ni trop long. Les domaines à rallonge ne sont pas très bien vus par le robot, et c'est pareil pour les domaines trop courts. Par exemple, « google.fr » est un nom de domaine trop court (6 caractères hors extension) alors que « siteduzero.com » (10 caractères hors extension) est bien meilleur.

On admet généralement que l'idéal serait d'avoir un nom de domaine **entre 10 et 15 caractères**. La limite valable serait de **20 caractères**, cela vous donne quand même une marge de manœuvre plutôt grande pour vous trouver un joli nom, cette information est néanmoins à prendre avec des pincettes dans le sens où les URL ne sont pas prises en compte par le robot. 😊

Construire des URL "SEO-Friendly"

Points fondamentaux de la construction d'URL !

De la précision dans votre arbo, divisez vos contenus !

Je vous l'ai déjà dit, la principale force d'un site web, c'est son contenu. Le problème, c'est que très vite, plus le nombre de pages web grandit plus on va avoir besoin de diviser son contenu dans des sous-répertoires. Un site qui compte plusieurs milliers de pages dans son patrimoine ne se divise pas de la même manière qu'un site qui en possède 20. Le premier des deux aura systématiquement besoin de passer par des sous-dossiers pour optimiser son référencement, le second, non.



J'ai un grand nombre de pages web sur mon site, mais moi j'aime bien que mes pages soient toutes à la racine de mon site, en quoi ça lui pose problème à Google ? C'est mon site après tout, je le divise comme je veux.

Pour répondre à cette question. Je vais ressortir la page de résultats de Google. Prenons encore une fois le cas de la recherche pour le cours de PHP du Site du Zéro.

► [Concevez votre site web avec PHP et MySQL - Le Site du Zéro](#) 
www.siteduzero.com/tutorial-3-14668-concevez-votre-site-web-av... - En cache
14 mai 2010 – Si vous avez aimé ce cours, vous pouvez retrouver le livre "Concevez votre site web avec PHP et MySQL" du même auteur, en vente sur le **Site du Zéro**, ...

Jetez un coup d'œil à l'URL en vert. Je vais maintenant vous la recopier ici, étant donné que sur Google, elle est tronquée.

Citation : Lien cours de PHP : Site du Zéro

siteduzero.com/tutoriel-3-14668-concevez-votre-site-web-avec-php-et-mysql.html

Cette URL est assez correcte, mais elle pourrait être optimisée, dans le sens où l'on aurait pu placer ce cours dans un sous-dossier, voire dans un sous-sous-dossier, afin d'avoir quelques mots-clés en plus que Google aurait considérés lors de son indexation. Le Site du Zéro possède quand même plusieurs milliers de pages, on pourrait tirer profit par exemple du nom des sections où les tutos sont classés.

Par exemple :

Citation : Lien cours de PHP optimisé: Site du Zéro

siteduzero.com/apprendre-php/tutoriel-3-14668-concevez-votre-site-web-avec-php-et-mysql.html

Tout de suite, c'est optimisé étant donné que l'on a un sous-dossier « apprendre-php » contenant déjà deux mots-clés dégageant l'intérêt de votre page : « apprendre » et « php ». Ces mots-clés, je les ai mis au hasard, on les travaillera vraiment une fois que nous serons plus avancés dans le cours. Comme d'habitude, tentez pour le moment de vous contenter uniquement d'une expression qui définit rapidement l'ensemble du contenu que l'on va trouver dans votre sous-dossier. Patience mes amis, patience.



Cet exemple pour le **Site du Zéro** n'est probablement pas le plus optimisé pour la structure du site. C'était uniquement un exemple, je suis persuadé qu'il faudrait quand même creuser un peu plus que cela pour une meilleure solution pour le cas du Site du Zéro. Comprenez juste que l'on doit être le plus descriptif possible ! Votre URL doit dégager à elle seule l'intérêt de votre page pour inciter le visiteur à cliquer dessus.

Vous avez probablement compris le concept maintenant ! Lorsqu'on a beaucoup de contenu sur son site, il faut diviser !



Nous découvrirons dans le prochain chapitre la notion de Google Juice ou Link Juice. Vous comprendrez alors à ce moment-là l'autre intérêt de diviser vraiment votre contenu dans des sous-dossiers.

Nommez correctement vos fichiers

Vos fichiers doivent absolument être nommés correctement, étant donné qu'ils apparaissent dans l'URL et donc dans les résultats de Google. Vous ne devez pas surcharger les noms de vos fichiers de caractères ou d'expressions inutiles. Nommez-les de la manière la plus simple possible, en favorisant l'intégration de mots-clés. Il faut éviter absolument tous les caractères inutiles dans les noms de vos fichiers. **Cela ne sert strictement à rien.**

Les tirets comme séparateurs

Dans vos URL, vous utiliserez les tirets « - » comme séparateurs plutôt que des underscores « _ » et ce, dans un souci d'uniformiser vos URL vis-à-vis de votre nom de domaine qui lui ne peut techniquement pas contenir d'underscores. Par ailleurs, retenez toujours que l'utilisation d'underscores dans les URL est une erreur fréquente qui coûte cher, et qui peut être très facilement évitée. Un fichier nommé « apprendre-php.html » est plus intéressant que « apprendre_php.html » dans le sens où l'URL sera uniformisée. En revanche pensez toujours au fait que les URL ne sont pas prises en compte par Google, vous n'aurez donc aucune pénalité en utilisant des *underscores*, c'est simplement « plus beau ».

Les caractères « complexes »

Les URL n'étant pas prises en compte depuis Google Panda, vous pouvez nommer vos fichiers comme vous le voulez, ce qui implique que vous pouvez y incorporer des accents si vous le souhaitez : prenons l'exemple de Wikipedia. Ces trois liens amènent vers le même article et deux de ces trois liens comportent des accents.

Citation : Lien 1° Guerre Mondiale - Wikipedia

http://fr.wikipedia.org/wiki/Premiere_Guerre_mondiale
http://fr.wikipedia.org/wiki/Première_Guerre_mondiale
http://fr.wikipedia.org/wiki/1ère_Guerre_mondiale

Wikipedia étant l'un des sites les mieux référencés au monde, c'est la preuve que Googlebot ne prend pas en considération les URL et que vous pouvez en faire ce que vous voulez ; celles de Wikipedia sont néanmoins très explicites pour les visiteurs, pourquoi ne pas utiliser des URL similaires à celle de Wikipedia ?



Voyez par ailleurs que Wikipedia possède de nombreux *underscores* dans ses URL, cela ne l'empêche pas d'être très visible sur Google !

Le H1 dans l'URL

La dernière partie de l'URL devra être votre titre `<h1>`. Cela permettra de faire apparaître votre `<h1>` dans la page de résultats de Google, et dans son URL : toujours bon pour inciter un visiteur à cliquer. 😊

Sous-domaine ou répertoire ?

Vous devez forcément vous demander quel type de découpage est le plus recommandé pour un site web : doit-on utiliser des sous-domaines, ou des répertoires sur le domaine principal ? Voici le format que vos URL prendront dans les deux cas, lorsque nous souhaitons accéder à une page de profil par exemple.

Avec sous-domaine	Avec répertoire
membre.votresite.com/profil.html	votresite.com/membre/profil.html

C'est simple : le sous-domaine viendra se greffer avant le nom de domaine avec un point (« . ») comme séparateur. Avant 2007, les sous-domaines s'employaient de manière totalement différente. En effet, pour Googlebot, un sous-domaine représentait carrément un nouveau site complètement différent du domaine principal, qui ne profitait donc pas de son référencement déjà construit.

Fin 2007, **Google a fait une mise à jour de son algorithme** : désormais, Googlebot considère un sous-domaine comme un répertoire du domaine principal. Il n'y a donc, aujourd'hui, que des différences très minimes entre utiliser un sous-domaine plutôt qu'un répertoire, et inversement.

Une autre règle d'or : un changement = une redirection.

Évitez les erreurs 404 !

C'est ce que trop de Webmasters ne font pas, et c'est également ce qui vous massacre un référencement acquis en plus de rapporter des pénalités. Changer l'emplacement d'un fichier ou le renommer entraîne obligatoirement un changement au niveau de l'URL d'accès de celui-ci. Seulement, Googlebot n'est pas devin. Au moment où il devra revenir voir votre page web, il empruntera le lien qu'il avait visité la dernière fois. **Et ce lien n'étant plus valide, le Googlebot tombera sur une belle erreur 404.**



Et croyez-moi, s'il y a bien une chose que Googlebot déteste par-dessus tout et qui vous rapportera énormément de pénalités : ce sont les erreurs 404.

Encaisser une erreur 404 est un **énorme point noir** pour son site. C'est pourquoi la toute première chose à faire sera de créer une page de 404 personnalisée, sur laquelle tous les utilisateurs, dont le Googlebot, seront redirigés dès qu'une 404 sera levée. Cela limitera déjà la casse. Créez un fichier `.htaccess` à la racine de votre site web, et redirigez les erreurs 404 sur un fichier d'erreur personnalisé, que l'on va appeler par exemple `"erreur.html"`. Les pages d'erreurs se gèrent en `.htaccess` grâce à l'aide de la directive **ErrorDocument**.

Code : Apache

```
ErrorDocument 404 http://www.site.com/erreur.html
```

Vous pouvez maintenant créer un fichier `erreur.html` le placer à la racine de votre site pour personnaliser une erreur ; c'est d'ailleurs le type de procédé qu'emploie le Site du Zéro. [Essayez vous-même.](#) 😊

Maintenant, cela ne suffit pas. Si l'internaute ou le Googlebot emprunte l'ancien lien de votre page et que celle-ci se trouve à un autre endroit, vous aimeriez naturellement que celui-ci soit redirigé sur la nouvelle page plutôt que sur la page d'erreur. En effet, Googlebot aura à cœur de visiter plusieurs fois la même page web dans le temps pour voir si celle-ci a été modifiée ou non et aller toujours plus loin dans son indexation. Ainsi, pour qu'il ait toujours le fil, **il va falloir faire une redirection**. Pour cela, nous allons encore une fois devoir passer sur le fichier `.htaccess` et travailler dessus.

Ouvrez-le simplement, et ajoutez une **redirection permanente** à l'aide de la directive `RedirectPermanent`.

Code : Apache

```
RedirectPermanent /ancienne-url.html http://www.site.com/nouvelle-  
url.html
```

Grâce à ce bout de code placé dans votre fichier `.htaccess` toutes les requêtes demandant le fichier `ancienne-url.html` seront automatiquement redirigées vers `nouvelle-url.html`. Ainsi, votre URL aura été déplacée correctement, Google la prendra à son tour en compte, et comme ça, vous n'encaisserez pas de pénalités.



Si vous êtes amenés à changer le nom de domaine de votre site, sachez qu'il est possible de le signaler à Google d'une autre méthode via les Webmasters Tools. Nous verrons cela en temps voulu. 😊

Comme il existe une infinité de manières de coder un site, adaptez ces solutions pour les outils avec lesquels est codé votre site, bonne chance. 😊

Précisions sur les redirections

Redirection 301

Petit quart d'heure technique : les serveurs et les clients, sur le web, communiquent grâce au protocole HTTP, qui n'est rien d'autre qu'une langue commune pour s'échanger des pages web, c'est pourquoi toutes les URL commencent par « `http://` ». La redirection que nous venons de voir est en fait **une redirection 301**, qui est tout bêtement une redirection exprimée en HTTP. C'est donc le serveur lui-même qui effectue cette redirection. Il s'agit-là de la manière la plus naturelle de rediriger des visiteurs ou robots vers une ressource qui a été déplacée. Le nombre 301 est un code qui correspond en fait à « redirection permanente (définitive) de ce contenu ». Les visiteurs et robots seront donc redirigés directement par le serveur lui-même vers le nouvel emplacement de votre page.



Il existe bien sûr d'autres codes en HTTP qui correspondent à différentes choses. Par exemple, le code 404 correspond à une erreur déclenchée lorsque la page n'a pas été trouvée. 😊

Ainsi, utiliser une redirection 301, c'est s'assurer de rediriger les visiteurs et les robots de la manière la plus naturelle possible vers la nouvelle ressource ! Nous n'aurions pas pu procéder avec du JavaScript par exemple, qui est une méthode beaucoup plus exotique et "sale".

Créer des redirections 301 sous Apache, NGinx

Comme nous l'avons vu, sous Apache, il suffit de créer un fichier `.htaccess` et d'y utiliser la directive `RedirectPermanent`.

Code : Apache

```
RedirectPermanent /ancienne-url.html http://www.site.com/nouvelle-  
url.html
```

D'autres serveurs HTTP existent, tels que NGinx, qui est lui aussi très utilisé. Cependant, j'ai une mauvaise nouvelle : les fichiers `.htaccess` n'existent pas sous NGinx, et les redirections 301 se réalisent différemment. Il vous faut donc ouvrir le fichier

/etc/nginx/nginx.conf, et y ajouter les liens qui suivent dans la section server{ }.

Code : Apache

```
location /ancienne-url.html {
    rewrite /chemin/de/votre/ancienne/page.html
    http://www.site.com/nouvelle-url.html permanent;
}
```

Autres redirections

Sachez que les redirections 301 ne sont pas les seules redirections existantes. Il en existe d'autres en HTTP. En fait, tous les codes de redirection commençant par le chiffre 3 sont des redirections. En voici quelques unes pour votre culture..

Code et nom	Effet
301 Moved Permanently	Redirection que l'on utilise pour indiquer qu'une ressource a été déplacée de manière permanente. C'est celle que l'on a utilisé.
302 Found	Ce code indique au client que la ressource demandée a été temporairement déplacée vers un nouvel endroit.
303 See Other	Le code 303 n'indique pas que la ressource demandée a été déplacée, mais elle renvoie en fait une autre adresse où la ressource peut être trouvée.
307 Temporary Redirect	Le code 307 indique quant à lui que la ressource demandée est temporairement disponible à une autre adresse, elle indique également l'adresse en question qui est retournée au client. Le code 307 est très semblable au code 302, car il a été fait à partir du code 302 afin d'en faire une version moins ambiguë.

Problématique des URL dynamiques

Les URL, en particulier celles qui sont générées directement par des langages tels que PHP posent rapidement un problème : elles sont très laides pour l'internaute qui ne peut pas savoir ce qu'elles contiennent.

Un problème d'URL

Des URL générées dynamiquement

Ces URL laides gérées dynamiquement possèdent des variables dans leur construction. Voici par exemple une URL qui aurait pu être générée par PHP.

Citation : URL par PHP

```
http://www.site.com/page.php?id=1&ion=le-site&titre=ouverture-!
```

Ce type d'URL est assez médiocre vous ne trouvez pas ? Pour le Googlebot pas de problème, il saura l'indexer, mais pour votre visiteur, vous ne trouvez pas que c'est dommage de le laisser se bagarrer avec beaucoup de variables pour savoir ce que va contenir la page web ? Il serait quand même bien de nettoyer un peu ces URL pour aider le visiteur à identifier le contenu de votre page à partir de l'URL. Dans le même style, voici un exemple d'URL encore plus mauvais.

Citation : URL très mauvaise

```
http://www.site.com/page.php?id=1
```

Ou encore pire (oui 🤪), une URL comportant la référence d'un produit que l'on souhaite vendre.

Citation : URL très très mauvaise

```
http://www.site.com/page.php?ref=KHDH0184DAA X578451F
```

Vous vous rendez compte ? Avez-vous une seule idée de ce que va contenir la page à partir de la référence du produit ? Non, et

c'est normal. Par ailleurs, il me semble correct de vous prouver à quel point il est désagréable pour un visiteur de ne pas pouvoir identifier l'intérêt d'une page web à partir de son URL. Voici le lien vers une vidéo YouTube :

<http://www.youtube.com/watch?v=7Hk5uVv8JpM>. Il est quand même difficile de savoir que la vidéo est une conférence sur le référencement animée par Google. Pourtant, je suis sûr que vous auriez cliqué dessus si cela avait été indiqué dans l'URL, avouez ! 😞

Aujourd'hui, nous avons de la chance, les URL ne sont pas prises compte dans l'indexation par Google. Il n'est donc pas si dramatique que ça d'avoir des URL laides. Cependant, vous devez savoir qu'à l'époque où Google accordait encore un impact relativement important aux URL, les sites qui proposaient des choses comme les références d'un produit en guise d'URL n'arrivaient jamais réellement à bien se positionner. Certains de ces sites faisaient l'effort de tenter de mettre à jour toutes leurs URL, mais il y avait un piège ! Certains oubliaient de faire des redirections 301, et forcément les 404 à répétition avaient tendance à ne pas trop plaire à Google.

Heureusement, cette époque est révolue et nous ne raisonnons maintenant plus qu'en terme de confort pour votre visiteur. Si votre site est aujourd'hui construit autour d'URL médiocres, vous pouvez le laisser vivre. N'oubliez pas que les URL ne servent aujourd'hui qu'à localiser des ressources (pages) sur le web. Si vous souhaitez néanmoins jouer avec le feu, il existe une solution plutôt sympathique mais assez complexe à mettre en oeuvre techniquement.

La solution

L'URL Rewriting

La solution à ce problème d'URL laides consiste en une technique que l'on met généralement en place via des fichiers .htaccess (dans le cas d'un serveur Apache) pour mettre au point des URL dynamiques mais propres. Cette technique se nomme l'« URL Rewriting », ou la « réécriture d'URL ». C'est quelque chose d'assez difficile à mettre en place étant donné que c'est très technique et que tout le monde ne sait pas forcément manipuler les fichiers .htaccess correctement. Grâce à l'URL Rewriting, on peut effectuer ce genre de modifications sur les URL.

Code : Autre

```
Une URL par PHP http://www.site.com/page.php?id=1&ion=le-site&titre=ouverture-!  
Même URL par PHP avec rewrite http://www.site.com/le-site/1-ouverture-!.html
```

Cette URL est quand même beaucoup plus propre en terme de confort pour votre visiteur, mais pas forcément. Le contenu est ainsi structuré et hiérarchisé. Vous devez maintenant savoir que ce problème d'URL propres existe et que c'est une plaie. Si vous avez compris, c'est déjà bien suffisant.

En résumé

- Votre nom de domaine est un choix très important.
- Optez pour un nom de domaine possédant un mot-clé en rapport avec votre thématique.
- Votre nom de domaine apparaîtra pour chaque page de votre site.
- Plus vous aurez de pages, plus vous devrez diviser votre arbo.
- Même si Google est maintenant tout à fait capable de traiter des URL contenant des paramètres, optez pour des réécritures propres.
- Les 404 sont très mal vus par Google. Un changement d'adresse = une redirection.

Un site web, une arborescence, des URL ! (2/2)

Vous allez voir dès à présent pourquoi il est important de hiérarchiser son contenu dans son arborescence. La raison que je vous ai donnée dans le chapitre précédent, c'était celle de la densité de mots-clés dans votre URL... seulement ce n'est pas tout ! Il y a aussi une autre cause à cela : le Google Juice (ou jus de lien) !

À défaut de se lancer dans l'agroalimentaire, Google nous offre encore tout un concept de référencement, que l'on va devoir décortiquer ensemble car le connaître sera l'une des clés du succès de votre référencement.

Le Link Juice ou Jus de lien !

Le Google Juice ! (ou Link Juice)

Un nouveau concept en SEO

Le Google Juice est vraiment un concept très important en SEO, et ne pas le connaître pourrait vous coûter quelques points en SEO. Heureusement, nous sommes ici pour l'apprendre, vous allez voir, ce n'est pas bien compliqué.



Avant de commencer à expliquer ce qu'est le Google Juice, vous devez savoir qu'un autre nom existe. On l'appelle le plus souvent "Link Juice", et sur les forums français, j'ai même vu l'expression "jus de lien", alors ne soyez pas surpris si vous voyez ces termes sur le web. Ils sont aussi très utilisés.

Maintenant cette petite note passée, répondons à la question que vous vous posez certainement depuis le début de ce chapitre :



Mais c'est quoi le Google Juice ?!

Pour bien comprendre le Google Juice, nous allons devoir comparer votre site web à un ensemble de gobelets dans lequel coule un liquide : le Google Juice. Chaque page de votre site constitue un petit gobelet différent d'un autre. Ainsi, vous avez autant de gobelets que vous avez de pages web sur votre site. Voici une image dans laquelle le gros gobelet qui se trouve tout en haut est la page d'accueil de votre site web. Essayez bien de projeter cette image sur votre site, je sais que ce n'est pas très évident.

Google



Maintenant, il faut savoir que ces gobelets ne vont pas rester vides très longtemps. Google va y verser du jus. Du Google Juice pour être exact. Cependant, Google a toujours le souci de la précision et il ne va pas le faire n'importe où. En fait, il ne va pas chercher bien loin et va tout de suite voir que le gobelet le plus élevé est votre page d'index : « / ». Ainsi, tout bêtement, Google va verser tout son jus de référencement ici, comme dans cette image.



C'est très bien. Mais nos autres gobelets sont un peu vides non ? C'est votre site, donc vos gobelets. Vous êtes donc libres de faire des petits trous dans votre page d'accueil, pour laisser le jus couler vers un autre gobelet. Et puisque c'est bien sûr ce qu'il faut faire on va jouer le jeu à fond et faire des trous un peu partout pour laisser le jus couler partout sur notre site.



L'idéal étant que chaque page web reçoive une partie du jus. L'un des gobelets a été volontairement mis à l'écart : aucun jus n'y coule pour vous montrer que c'est vous qui choisissez où vous voulez faire couler votre jus. 😊



Euh excuse moi, mais je n'ai rien compris. Tu ne veux pas plutôt me dire à quoi ça sert de m'avoir expliqué ça ?

Google va verser tout son jus dans votre page d'index. En gros : il va verser tout votre potentiel de référencement dans votre page d'index. Par la suite, ce potentiel, cela va être à vous de l'acheminer vers toutes vos autres pages web en perçant des petits trous dans votre page d'accueil pour laisser le jus tomber vers vos autres pages web. Au final, plus une page contiendra de jus, plus elle sera importante pour Google.

Si vous faites cela (et c'est bien sûr ce qu'il faut faire) votre page d'accueil va perdre un peu de Google Juice, mais au profit de vos autres pages ! Il n'y a pas le choix, il faudra faire des concessions ! Mais faites bien attention, car si votre page présente deux liens vers deux pages différentes, le Google Juice qui sera acheminé sera moins concentré pour chacune des pages. Et de même si elle présente 3, 4, 5... liens. Ainsi, les pages qui contiendront le plus de Google Juice et de manière la plus concentrée à la fin, seront les pages qui auront le meilleur potentiel de référencement et qui pèseront le plus lourd dans la balance lors de l'indexation par Google. C'est là que le terme de hiérarchiser son contenu prend tout son sens, et c'est également là qu'il va falloir penser à l'arborescence de son site. Comme je vous l'ai dit, un site qui possède 20 pages ne se divise pas de la même manière qu'un site qui en possède plusieurs milliers. Vous devez acheminer le jus de manière intelligente : à travers votre arborescence. C'est toujours vis à vis de ce jus que vous allez devoir penser à diviser vos contenus.

Acheminer le jus à travers son site

Transmettre le jus aux autres pages

Les liens hypertexte

Voilà maintenant plusieurs minutes que je vous parle sans arrêt de créer des trous dans des gobelets/pages web. Alors il y a forcément une question que vous vous posez depuis tout à l'heure :



Et comment est-ce que je peux acheminer le Google Juice à mes autres page ? Tu parles de percer des trous dans les pages depuis tout à l'heure ? Il y a une balise HTML `<hole>` ou quelque chose dans le genre ?

La réponse est non, bien sûr ! En fait, c'est tout bête. Pour percer un trou dans votre page web... il faut tout simplement utiliser un lien HTML avec la balise `<a>`. La page qui recevra le jus sera la page qui sera renseignée dans `href`. Par exemple si dans `index.html` je créer un lien `` la page `jerecoislejus.html` recevra du jus de la part de la page `index.html`.

Au final l'idéal serait qu'au moins un lien pointe sur chaque page de votre site pour l'ensemble de votre site web. Ainsi, vous ne prendrez aucun risque : tous vos pages recevront un peu de votre jus. Googlebot et vos visiteurs seront donc ainsi capables de trouver toutes les pages de votre site web. Dans le cas de Googlebot, c'est également comme cela qu'il va pouvoir indexer toutes vos pages, en lui donnant des liens vers ce que vous voulez qu'il indexe. N'oubliez pas son fonctionnement : il ouvre tous les liens d'une page, et il envoie le code HTML à Google si on ne l'en empêche pas.



Vous pouvez voir qu'aucun lien ne mène vers le gobelet tout en bas à droite. Maintenant je vous conseille de créer un lien vers ce gobelet, car encore une fois, au moins un lien doit mener vers chaque page de votre site. Pour la suite, sachez quand même que votre page d'index ne doit envoyer du jus qu'à vos sections majeures, et c'est à elles ensuite, d'acheminer ce jus vers les autres petites pages. Comme vous pouvez le voir dans cette image, il n'y a aucun Google Juice qui coule de la page d'index tout en haut vers les petites pages qui se trouvent au troisième rang. 😊



Une très mauvaise répartition du jus est de créer un lien vers chacune de ses pages directement sur la page d'index. C'est encore une fois une très grosse erreur à ne pas faire.

En effet, n'ayez pas la main trop lourde sur les liens. À vous de voir comment vous souhaitez acheminer votre jus, mais retenez bien qu'une page qui contient 3 000 liens ne sert à rien, puisque le potentiel de référencement devient trop divisé. Vos pages ne pèseront donc presque rien.



N'oubliez pas que l'on parle ici toujours de lien interne à son site. Lorsqu'on parlera de nos actions en externe, vous verrez que la création de lien et le Google Juice d'une manière générale prendra encore une autre dimension. Ce sera d'ailleurs l'occasion pour nous de parler du PageRank et des backlinks ! Vous verrez que l'on peut envoyer notre jus à d'autres sites, mais aussi en recevoir de leur part ! 🤖

Ne gaspillez pas bêtement votre jus !
Attention aux 404 !

Ne perdez pas votre jus

La première chose à laquelle il va falloir penser lorsque l'on achemine du jus, ce sont les erreurs 404. Vous vous souvenez très probablement qu'il n'y a pas longtemps, je vous disais qu'il ne fallait en aucun cas que Googlebot tombe sur un lien qui donne sur un 404. Si vous voulez connaître la raison pour laquelle je vous avais dit cela : c'est le Google Juice !

En effet, un 404 correspond à une perte de jus. **Le 404 not found** indique la page que vous avez demandé n'a pas été trouvée. Ainsi, cela correspond à faire un lien dans une page (un trou dans un gobelet) avec à l'arrivée aucune ressource (aucun gobelet où le jus arriverait). Le jus tomberait donc dans le vide et serait perdu. C'est donc du potentiel de référencement bêtement gaspillé. Voici en image ce que donne un 404.



Le jus tombe ici dans le vide, il

Ce jus, il ne faut justement le perdre en aucun cas. Voilà pourquoi je vous dis que la pire abomination pour votre site web serait d'avoir des 404 en pagaille. Je pense que l'on ne pourra jamais empêcher d'avoir des 404 sur son site, mais on peut toujours les limiter. Limiter comment ? En passant par une page 404 personnalisée comme nous l'avons vu précédemment. Bien sûr, l'idéal est de n'avoir aucun 404 levé sur son site : une page d'erreur personnalisée ne vous dispense en aucun cas de corriger vos liens morts. 😊

Boucher les trous avec rel="nofollow"

Tous les liens comptent pour Google : si un de vos membres crée un lien vers une page sur votre forum, eh bien du Google Juice va passer sur ce lien... Dans le cas où c'est un 404, vous voilà bien embêté.

La solution est de passer par un ``. En effet, le `rel="nofollow"` agit un peu comme un bouchon. Le `<a>` crée un trou puisqu'un lien est créé, mais il est immédiatement bouché par le `rel="nofollow"`. Ainsi, ceci permet à votre communauté d'utiliser des liens, mais sans agir sur votre Google Juice.



C'est bouché ! Le Google Juice ne

Vous gardez ainsi toujours la main sur votre référencement. Si vous ne voulez pas que vos membres aient la possibilité d'impacter sur votre Google Juice, n'hésitez pas à programmer quelque chose qui ajoutera un `rel="nofollow"` aux liens sur vos forum. Pour des sites où le contenu est généré à 100% par la communauté, le problème sera quand même plus délicat à traiter. À vous de trouver les solutions en fonction de votre site.



Ce dont vous devez vous souvenir, c'est que chaque lien crée un trou et le Google Juice s'échappera par ce trou. Mais que chaque `rel="nofollow"` bouchera le trou. Dans le cas des 404, vous n'êtes en aucun cas dispensés de les corriger.

Précisions sur le Google Juice

Quel jus est transmis ?

Le pourcentage de jus

Il est très difficile de quantifier le Google Juice qu'une page détient, c'est pourquoi au lieu de nous demander « Quelle quantité de jus est transmis ? », nous allons plutôt nous demander « Quel pourcentage de jus est transmis ? ». À la base, une page d'index contient 100% de votre jus, puisque c'est à cet endroit que Google va verser tout le potentiel de référencement de votre site. Nous allons reprendre nos petits gobelets qui symbolisent nos pages web.

Transmettre tout son jus

Pour transmettre la totalité c'est très simple, vous devez tout simplement créer un lien, et un seul lien d'une page A vers une page B. La page A transmet ainsi 100% de son potentiel de référencement, jus est le plus concentré possible.



Fractionner son jus

Si vous créez deux liens amenant vers deux pages différentes (d'une page A vers les pages B et C), vous allez fractionner votre jus en deux. Ainsi, la page A va transmettre 50% de son potentiel de référencement aux pages B et C. Pour fractionner encore plus votre jus, c'est très simple, vous devez simplement diviser le jus transmis par le nombre de liens présents sur la page. Si vous avez 10 liens présents sur une page A qui pointe vers 10 pages différentes, alors, ces pages recevront chacune 10% du jus de la page A. C'est simple et logique.



Cas particulier

Cas particulier du rel="nofollow"

Le `rel="nofollow"` constitue un cas particulier que vous devez connaître. Plus haut, je vous ai dit que vous deviez utiliser un `rel="nofollow"` pour « boucher » les trous. Vous empêchez ainsi le jus de passer. Cette technique est donc intéressante à utiliser dans plusieurs cas de figure. Je citais tout à l'heure l'exemple des membres qui créent des liens sur les messages des forums : la solution est d'ajouter de manière automatique un `rel="nofollow"` afin que le jus reste concentré sur les pages de votre forum, ainsi, vos membres ne pourront pas impacter sur votre manière de gérer votre Google Juice. C'est intéressant me direz-vous, mais il faut également savoir ce que va faire le `rel="nofollow"` si vous souhaitez créer d'autres liens. Admettons que vous possédiez une page A qui comprend deux liens.

- Le premier lien est en `rel="nofollow"` et pointe vers une page B ;
- Le second lien est un lien courant et pointe vers une page C.

Nous avons donc deux possibilités et il nous appartient d'en exclure une. Soit la page C reçoit 100% du jus de la page A. Soit la page C reçoit 50% de la page B, et le reste ne sert pas. La bonne solution est la deuxième.



En effet, un `rel="nofollow"` ne signifie pas que le jus sera réinvesti dans les autres liens présents sur la page, cela signifie simplement que le lien sur lequel le `rel="nofollow"` est placé ne transmettra pas de jus. Saisissez bien cette nuance, car c'est essentiel pour bien comprendre l'action du `rel="nofollow"`. Si vous souhaitez transmettre un jus le plus concentré possible, le mieux est de ne pas créer de liens à tout va : chaque balise `<a>`, qu'elle soit en `rel="nofollow"` ou non divisera le jus transmis par la page.



Par le passé, le jus des liens en `rel="nofollow"` était réinvesti dans les autres liens de la page. Comme nous l'avons vu dans ce chapitre, ce n'est plus le cas aujourd'hui, et ce depuis deux ans. Faites donc attention en créant des liens sur vos pages !

En résumé

- Le Google Juice est aussi appelé Link Juice sur le web
- Il correspond à un jus que Google fait couler sur votre page d'accueil.
- Vous pouvez (et devez !) transmettre ce jus aux pages de votre site
- Ne perdez jamais votre jus avec des 404. C'est la pire erreur que vous pouvez faire avec le Google Juice.
- En principe chaque page web de votre site doit être ciblée par un lien quelque part en interne.

L'importance des mots-clés !

Depuis le début de ce cours nous apprenons les bases du référencement « *in-page* » : l'importance du contenu, du code HTML, de l'arborescence, des liens entre les pages... Nous allons maintenant apprendre un nouveau concept clé du référencement : les mots-clés.

Une question de mots-clés !

Un mot-clé, des mots-clés

L'ère du contenu

Je vous le dis depuis le début du cours, avec Google, nous sommes passés à l'ère du contenu : cela signifie que le contenu de votre page web est considéré comme son intérêt principal, c'est donc l'élément numéro 1 de votre page... et de votre site web par extension.

Dans ce contenu - généralement textuel - nous pouvons identifier des éléments plus importants. Nous l'avons vu il y a quelques chapitres, les balises telles que `` nous permettent de les faire ressortir et de les mettre bien en évidence pour le robot. Seulement, qu'est-ce que tout cela signifie dans le monde du référencement ? C'est très simple : lorsque vous allez lancer votre site web et commencer à produire du contenu, vous allez devoir définir une liste de mots-clés que votre site va utiliser.



Le fonctionnement des moteurs de recherche et les mots-clés

Vous utilisez certainement Google tous les jours ou presque, donc vous connaissez son fonctionnement de base. Vous tapez une recherche sur Google, et celui-ci vous renvoie des résultats. Ce que vous tapez dans la barre de recherche, c'est ce qu'on appelle des mots-clés. À partir de ces mots-clés que vous avez tapés, Google est capable de vous renvoyer une liste des sites utilisant ces mots-clés. Il va donc falloir que vous établissiez une liste des mots-clés que l'on pourrait potentiellement taper sur Google pour trouver votre site, et qui vont donc le caractériser.

Il ne faut franchement pas choisir ses mots-clés n'importe comment, et y réfléchir un minimum... si vous faites un site web sur Mario Kart, vous n'aimeriez pas qu'on le trouve sur Google en recherchant des pièces de voiture d'occasion non ? 🤔 Je force bien sûr le trait pour vous aider à comprendre, mais c'est l'idée. Nous allons donc devoir établir une liste de mots-clés pour votre site, avant de le lancer sur le web.

Établir une liste de mots-clés pour son site

Deux facteurs importants

Il va falloir bien réfléchir aux mots-clés que nous souhaitons pour votre site : un bon mot-clé est un mot-clé qui correspond à deux facteurs.

- **Il doit être beaucoup recherché sur Google :**
Un bon mot-clé doit être beaucoup recherché sur Google par les internautes. En effet, on crée un site pour qu'il soit visité : si vous choisissez des mots-clés que personne ne tape, votre site risque de ne jamais décoller malheureusement.
- **Il doit bien correspondre à votre site :**
Un bon mot-clé doit correspondre à votre site : si les mots-clés que vous choisissez ne correspondent pas à votre activité, vos visiteurs risquent de ne pas rester très longtemps sur votre site.

Commencez par aller chercher une feuille de papier... et dressez une liste de vos mots-clés. C'est assez simple, imaginez ce que vous voudriez que l'on tape sur Google pour trouver votre site web, et notez cela sur votre liste. Demandez à vos amis, à vos visiteurs actuels (si votre site existe déjà), bref ! Réfléchissez un maximum, le tout ici est d'arriver à une liste complète des mots-clés qui caractérisent votre site.

Si vous effectuez ce travail correctement et que vous êtes inspirés, vous devriez arriver à une liste assez complète de mots-clés. Tentez maintenant de les analyser vous-même : quels sont les mots-clés les plus importants, ceux qui présentent le plus de potentiel... Rangez vos mots-clés selon leur importance, le tout est d'arriver à hiérarchiser ces mots-clés de sorte à distinguer ceux qui seront au premier plan, donc très utilisés sur votre site, et les autres moins intéressants.

La création de votre site

Utilisez vos mots-clés !

Vous avez établi une liste des mots-clés et vous en êtes satisfait ? C'est très bien, vraiment... mais maintenant il va falloir les incorporer à votre site ! Je vous assure que vous savez le faire, il suffit simplement de les incorporer dans toute la structure de votre site : c'est ce que l'on a appris lors des chapitres précédents ! 🤖

Les zones suivantes devront être riches en mots-clés :

- Vos balises `<title>`
- Vos liens hypertextes `mots-clés`
- Balise `<h1>`
- Balise `<h2>` à `<h6>`

Et si le tout se trouve dans les 600 premiers pixels de vos pages web, c'est parfait. Bien sûr, vous devrez incorporer vos mots-clés partout où vous avez la possibilité d'inclure du texte. Les balises `` et `` sont là pour vous aider à faire ressortir des mots-clés dans votre contenu.



Bien sûr, lorsque je dis « incorporer » vos mots-clés, il ne s'agit pas de les intégrer sous une forme de liste à vos pages web... ils doivent s'intégrer naturellement à votre texte et le visiteur n'est pas censé les remarquer au cours de sa lecture. C'est facile si vous avez choisi des mots-clés en rapport avec votre site comme je vous l'ai expliqué.

Le suivi des mots-clés

Utiliser un outil Analytics

Personne n'est capable aujourd'hui de vous dire quel mot-clé peut vous rapporter quelles visites tant que vous ne les avez pas testés vous-même. Oui ! Il va falloir faire un suivi, par exemple mensuel. Lorsque vous aurez choisi certains mots-clés et que ceux-ci ne rapportent rien comme visites, il faudra peut-être songer à les changer sur une prochaine période. Pour cela, il vous faudra utiliser un outil « Analytics » comme Google Analytics.



C'est un service web d'aide aux webmasters. Concrètement, vous aurez à inclure un petit code JavaScript dans vos pages web afin d'obtenir des rapports très détaillés de la part de Google. Vous saurez ainsi quelles recherches sur Google ont amené quel visiteur à visiter votre page... il faudra utiliser ces nombres et statistiques afin de peaufiner toujours un peu plus le choix de vos mots-clés ! Google Analytics n'a pas le monopole sur les outils Analytics, il a également un bon concurrent appelé XiTi. Vous pourrez bien sûr utiliser les deux simultanément, c'est ce que fait le Site du Zéro.



J'ai encore beaucoup de choses à vous dire sur les mots-clés et le référencement en interne en général. Pour ne pas faire trop d'informations d'un coup, nous commencerons à utiliser des outils Analytics bien plus tard dans ce cours...

Choisissez vos mots-clés grâce à Google !

Utilisez Google Insights

Les tendances des recherches

Vous n'êtes pas convaincus par le choix de vos mots-clés ? Tant mieux, on a un moyen de faire les choses beaucoup plus proprement. Je vous expliquais plus tôt qu'un bon mot-clé se devait de respecter deux principes : il doit être souvent recherché sur Google et correspondre à votre site. Vous êtes le seul à pouvoir choisir des mots-clés qui correspondent à votre site. Personne ne peut vous aider pour ceci. En revanche, on peut obtenir de l'aide directement de la part de Google pour trouver les mots-clés qui sont beaucoup tapés sur le moteur de recherche.

Google a eu l'idée de rendre ces informations publiques par le biais du service Google Insights. Nous allons ainsi savoir ce qui est tapé sur Google tous les jours !



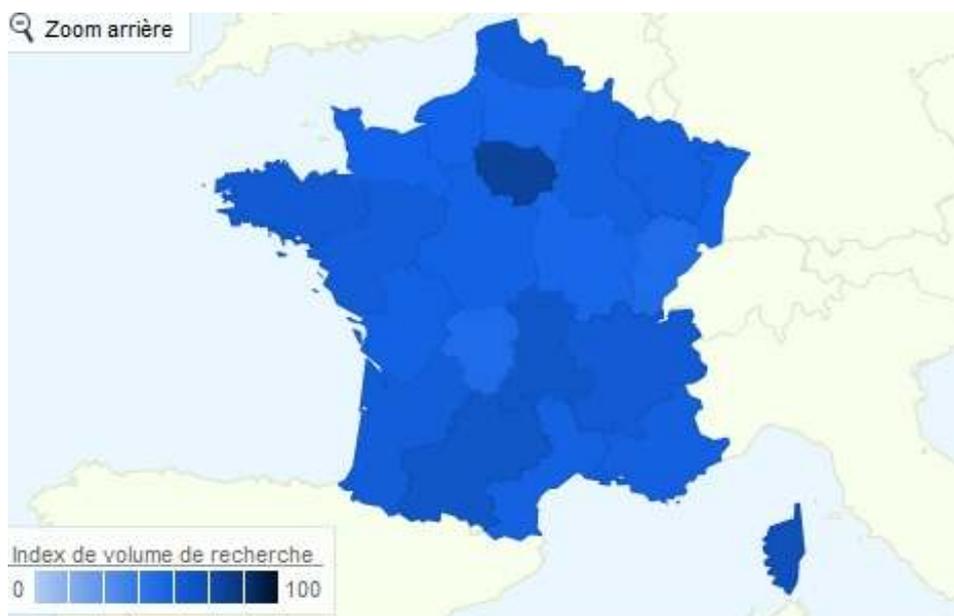
Google Insights n'est PAS Google Analytics. L'un sert à analyser les tendances des recherches, l'autre à analyser le trafic d'un site et sa provenance.

Accédez à Google Insights

Google Insights est accessible sur une URL publique et vous n'avez pas besoin d'avoir de compte Google afin de l'utiliser. Tout le monde peut exploiter la puissance de cet outil, néanmoins, si vous l'utilisez en étant connecté à un compte Google, vous obtenez la possibilité d'avoir les données statistiques sous formes de chiffre, et vous pourrez les télécharger directement au format .csv : je vous encourage donc à utiliser Google Insights avec un compte Google actif. Rendez-vous sur [Google Insights](#).

Analysons les tendances de recherche

Nous allons ici nous mettre dans la peau d'un internaute qui s'intéresse à l'informatique. Nous allons donc taper simplement « informatique » dans la barre de recherche de Google Insights. Le terme « informatique » étant purement français, nous allons devoir recadrer cette recherche sur la France. C'est très facile, utilisez le formulaire ! Première chose agréable à constater, Google Insights est capable de nous renvoyer un classement des régions françaises qui tapent « informatique » sur Google.



Il semblerait que l'on tape plus «

informatique » qu'ailleurs en Ile de France

Ce n'est pas tout, puisque l'on peut voir aussi quelles sont les recherches Google dans lesquelles le mot-clé « informatique » est présent : on peut voir qu'en France, c'est de se former à l'informatique qui intéresse surtout !



C'est la formation qui intéresse le

plus !



Notez que toutes les statistiques sont normées afin de faciliter les comparaisons.

Je pense que vous n'avez pas plus besoin de moi pour appréhender cet outil tant il est simple d'utilisation ! Tapez simplement les mots-clés que vous voudriez utiliser sur votre site, et voyez leur densité d'utilisation. Vous pourrez également voir que Google Insights propose des courbes pour évaluer cette densité dans le temps... bref, tout est très complet et devrait beaucoup vous aider.

Bien sûr, n'oubliez jamais de recadrer sur les régions du monde qui utilisent la langue de votre site. Rechercher un mot-clé français au Japon n'a que peu de sens, sauf si vous ciblez des visiteurs japonais francophones. 😊

Le concept de la longue traîne

Expliquons la longue traîne

Origine du terme

La longue traîne est à la base un concept qui n'avait rien à voir avec le SEO. La première personne à avoir employé cette expression est un journaliste nommé Chris Anderson, il tentait d'expliquer l'activité du site Amazon.com, c'est alors qu'il employa l'expression « Long tail » pour désigner les ventes d'Amazon.

Il s'était rendu compte d'une chose très intéressante vis à vis des activités du site. Amazon génère son chiffre d'affaire en vendant des produits. Chris Anderson a remarqué que l'on pouvait classer ces produits en deux catégories :

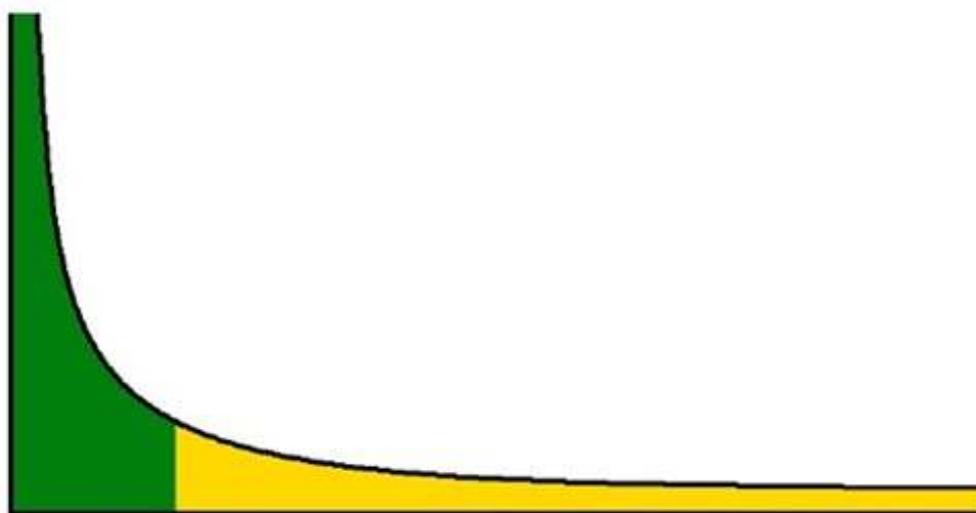
- **Des produits très vendus**

Il y avait des références qui étaient très vendues. En effet, Amazon comportait quelques produits, qui se vendaient par centaines. Imaginons deux ou trois références constamment en rupture de stock car elles se vendent énormément. Ces ventes constituaient 20% du chiffre d'affaire d'Amazon.

- **Beaucoup de produits très peu vendus**

À côté de cela, il y avait tous les autres produits. Des milliers de références qui ne se vendaient que très peu... pas plus de deux ou trois exemplaires par référence. Ces ventes constituaient 80% du chiffre d'affaire d'Amazon.

Vous devez penser que je me suis emmêlé les pinceaux, mais non ! C'est tout-à-fait logique : les produits qui se vendent en très petit nombre sont nettement plus nombreux que les produits qui se vendent en grand nombre. Si on les met tous ensemble, les produits qui se vendent en très petit nombre rapportent nettement plus d'argent que les « best-sellers » ! C'est ce que l'on appelle la longue traîne ! Un petit schéma simplifié ne serait pas de trop.



Vous pouvez voir que la partie verte qui représente les produits très vendus ne représente qu'une toute petite partie du chiffre d'affaire. En revanche la longue traîne en jaune ne monte pas très haut puisque ces produits se vendent peu, mais elle est nettement plus étendue ! C'était tout ce qu'il y avait à comprendre : les produits qui se vendent peu mais qui sont présents en grand nombre rapportent énormément par rapport aux articles qui se vendent beaucoup mais qui sont présents en très petit nombre !

Quel rapport avec les mots-clés ?

C'est très simple, la longue traîne existe aussi sur Google ! Imaginez que vos références (produits) soient vos mots-clés. Vous devrez choisir des mots-clés qui généreront beaucoup de visiteurs, mais sans négliger la longue traîne. Pour le SEO : la longue traîne est un très gros panel de mots-clés, donc chacun vous rapportera un nombre très limité de visiteur, mais bout à bout, c'est ce sont eux qui rapporteront presque la totalité de votre trafic. Pour cela, il va falloir avoir beaucoup de mots-clés.

Les mots-clés best-sellers

Ce sont à coup sûr ceux qui caractérisent le mieux votre site. Pour le Site du Zéro, il pourrait s'agir de « tutoriel » et « informatique ». Ce sont des mots-clés qui seront beaucoup tapés sur Google et qui vous rapporteront un très grand nombre de visites ; ce sont à coup sûr ces mots-clés que vous avez écrit sur votre liste tout à l'heure. Pour Amazon, ils correspondent aux articles qui étaient présents en petit nombre mais qui se vendaient énormément.

Les mots-clés de la longue traîne

Pour enclencher l'effet longue traîne sur votre site, la solution est simple, vous avez juste à produire beaucoup de contenus ! Avec 20 000 pages web sur votre site web, qui seront toutes travaillées comme on l'a appris précédemment, Google pourra analyser 20 000 codes HTML différents, avec 20 000 possibilités de mots-clés différents dans les balises `<title>`, `<h1>`, le contenu... C'est là que l'effet longue traîne prend tout son sens ! Si vous êtes 20 000 fois présent sur Google et que chacune de vos pages génère 2 ou 3 visites dans la journée avec vos mots-clés de la longue traîne : croyez-moi que vous ferez des bonds en voyant vos statistiques, et on ne parle même pas des « best-sellers » encore !

En plus, vous savez déjà que Google est un amoureux du contenu, c'est pour lui l'élément le plus important d'un site web. Plus vous en avez mieux c'est. Plus le Googlebot doit récupérer du HTML, plus il va aimer votre site et dès lors, il n'hésitera pas à booster votre référencement... alors créez beaucoup de contenu pertinent et « SEO-Friendly », et croyez bien qu'au fil du temps, votre site entrera dans un cercle vertueux, le tout est d'arriver à insuffler une bonne dynamique. Une fois la machine lancée, elle ne s'arrête plus !

En résumé

- Choisir de bons mots-clés est capital pour réussir son référencement, créez une liste et aidez-vous de Google Insights.
- Intégrez vos mots-clés de manière naturelle à vos pages web.
- Effectuez un suivi de vos mots-clés pour analyser le succès de votre référencement.
- Produisez beaucoup de contenu pour que la longue traîne s'enclenche.

La sémantique et le référencement

Le pari de l'avenir pour le référencement et pour le web en général est la « sémantique ». Vous devez probablement avoir déjà vu ce mot quelque part, tout comme « Web 3.0 ». Il est l'heure de dégrossir ces concepts un peu étranges mais qui sont très intéressants.

La sémantique et le référencement Le pari de la sémantique

L'histoire du web

Le web est âgé d'une vingtaine d'année tout au plus, et depuis sa création, il n'a cessé d'évoluer à une vitesse folle. Pour s'y retrouver, on s'est mis à vouloir distinguer les grandes périodes de l'histoire du web par des « numéros de version ». Vous avez donc forcément entendu parler de « Web 2.0 » ou de « Web communautaire ». Voyons un peu de quelle manière le web a évolué ces dernières années.

Le Web 1.0 dit « statique » - Années 1990

À ses débuts, le web était quelque chose de très simple. Avec le recul, on a appelé cette période le « Web 1.0 ». C'était une période où les sites web étaient statiques. Il n'y avait aucun échange entre les personnes : consulter un site web se résumait alors à lire une feuille de papier numérisée, ni plus ni moins. Le Webmaster du site proposait un contenu, et le visiteur le lisait. De là, il ne pouvait se passer aucune autre interaction entre le visiteur, le site, et le Webmaster. En bref, **le visiteur n'était que spectateur**.

Dans l'absolu, vous pouvez encore créer des sites type Web 1.0, il vous suffit simplement de créer des documents HTML et de les envoyer sur votre serveur. Bon cela marche, mais autant vous dire que les choses pourraient quand même être plus intéressantes que cela.

Le Web 2.0 dit « communautaire » - Fin des années 1990 et années 2000

La fin des années 1990 correspond à une énorme évolution dans le web. En effet, nous sommes passés à ce moment-là au « Web 2.0 », un web communautaire où le visiteur est maintenant acteur. C'est pendant cette période que des choses qui ne nous quittent plus aujourd'hui sont apparues.

- Les flux RSS
- Les réseaux sociaux
- Les sites de partage

Chaque visiteur est désormais acteur du web, on lui propose des solutions pour interagir très facilement avec les autres visiteurs, et tout le monde se sent bien en communauté en participant à la vie du site. Le Site du Zéro est typiquement un site « Web 2.0 ».

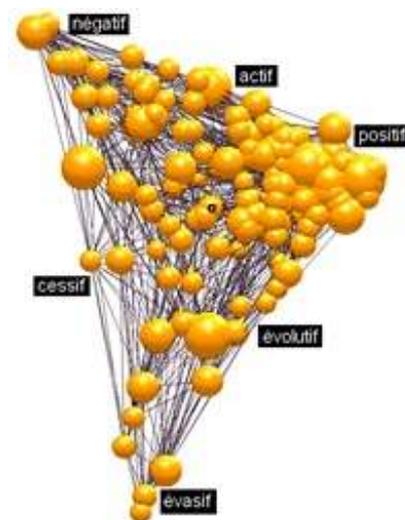
Avec le web 2.0, un nouveau problème se pose. Puisque tout le monde sur le web devient acteur et peut produire du contenu, il est dans le même temps de plus en plus difficile de suivre l'évolution d'un thème ou d'une information sur le web. En effet, tout le monde crée du contenu via des forums, des blogs, des réseaux sociaux... le volume d'information avec lequel il faut jongler a donc considérablement augmenté. On fait donc actuellement évoluer le web vers une nouvelle version, le « web 3.0 ».

Le Web 3.0 dit « sémantique » - De nos jours...

Il existe un débat concernant si oui ou non nous sommes aujourd'hui dans le web sémantique. Certains disent que oui, certains disent que non, et certains disent qu'on est entre le web communautaire et le web sémantique, autant vous dire que c'est un peu compliqué et que personne ne risque vraiment de tomber d'accord. Néanmoins, le web sémantique reste le grand pari du W3C pour les années à venir, Alors concrètement le web sémantique, c'est **quoi** ?

Le web d'aujourd'hui n'est rien d'autre qu'un ensemble de pages HTML, dont les moteurs de recherche arrivent à définir le thème par une suite de mots-clés présents. Par exemple, on arrive à définir le Site du Zéro comme un site de tutoriels, parce que le terme « tutoriel » est présent un peu partout sur le site. Le problème, c'est que si un humain sait ce qu'est un tutoriel, une machine, elle, ne sait absolument pas ce que c'est.

L'idée est donc de permettre l'utilisation de nouvelles balises HTML (notamment avec HTML5) pour permettre de aux internautes de ne plus créer de simples informations textuelles sur le web, mais **des données qui ont du sens** pour une machine. Prenons un



exemple apparu avec HTML5. Sémantiquement, pour définir un menu en XHTML, vous passiez par par exemple par une div portant l'id menu : <div id="menu">. Aujourd'hui, en HTML5, on a la balise <nav>. Concrètement, les machines comprennent que <nav> est un élément de navigation sur le site, alors que la balise <div> n'est rien d'autre... qu'un élément HTML block !

Dans le web 3.0 (sémantique), on va quand même un tout petit peu plus loin que cela, car on va avoir suffisamment de balises pour couvrir tous les types de données que l'on peut trouver sur le web. Voilà un petit extrait, vous imaginez que la liste est vraiment très longue..:

- Une localisation géographique
- Une carte de visite
- Un CV
- Un évènement dans le temps
- ...

Vous ne comprenez pas trop l'utilité du web sémantique ? Normal, j'ai omis de vous parler d'un facteur très important ! Sachez que de nouveaux langages sont envisagés par le W3C. Ces langages seront adoptés comme standards au même titre que le HTML par exemple. Parmi ces langages, nous aurons une sorte de langage de requête très puissant (comme le SQL !) pour obtenir des données présentes sur les sites web... et les utiliser dans des programmes ! Imaginez le potentiel de ces technologies à venir, il est quasiment illimité. 😊

Bien sûr, nous ne sommes pas totalement entrés dans le web sémantique, mais il faut comprendre que nous arriverons dans une situation où le web sera devenu intelligent, les machines seront capable de **comprendre** les données, au lieu de ne voir que du texte. Nous obtiendrons les informations plus facilement que jamais : c'est le web qui viendra à nous !



Le W3C est en pleine phase de développement de ces nouveaux standards... oui, cela représente une quantité de travail assez phénoménale, et on ne parlera réellement de web sémantique que dans une dizaine d'années au moins !

Il est difficile d'imaginer le futur, mais imaginons le cas d'un moteur de recherche comme Google : une recherche sur le web sera centralisée, une seule recherche n'enverrait plus une liste de sites non-ordonnée, mais plutôt un ensemble de données exhaustif et centralisé, car sur le web sémantique, tout est cohérent. Une seule recherche me permettrait donc de tout savoir sur les oiseaux par exemple, alors qu'en 2012, je dois rechercher sur plusieurs sites de passionnés d'ornithologie pour obtenir toutes les informations qu'il me faut.

Les moteurs de recherche intègrent de plus en plus les technologies qui feront le web sémantique, et nous disposons déjà de « balises HTML » plus évoluées que les balises courantes. Ces balises nous permettent réellement de définir des données sur nos pages web, données que Google exploite déjà en partie... c'est donc l'occasion pour nous d'apprendre à définir des données dans nos pages web, car cela aura un réellement impact sur le référencement. Dans la suite de ce chapitre, nous prendrons l'exemple de l'auteur d'un blog appelé Jacques. Nous définirons les informations de ce monsieur comme des données du type "Personne" grâce aux balises spéciales que j'ai évoqué. 😊



On parle également de « Web 4.0 », et certaines personnes pensent que le Web 4.0 va arriver avant le Web 3.0. Enfin, tout cela est bien compliqué, et nous allons rester centrés sur la sémantique, sinon cela va devenir vraiment complexe.

Le référencement à travers l'évolution du web

Le référencement évolue à travers le web

Le référencement du Web 1.0

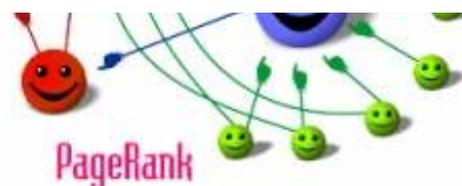
Au fur et à mesure que le web a évolué, la manière de se référencer a aussi évolué. En effet, le référencement est toujours le reflet du web. Cela implique que le passage au web sémantique va aussi impliquer de gros changements dans votre façon de concevoir votre site web. Le Web 1.0 était un web statique. Le visiteur n'était pas acteur du site, mais uniquement spectateur. Ainsi, le référencement était aussi très centré sur le site en lui-même, il se n'agissait que de placer les bons mots-clés en bon nombre dans les balises <meta /> des pages web.

Le référencement du Web 2.0

Avec le Web 2.0 tout a changé. Le web est devenu beaucoup plus communautaire, beaucoup plus ouvert, et beaucoup plus centré sur la communauté. Le référencement est aussi calqué sur ceci : vous savez maintenant que vous ne pouvez pas faire référencer votre site sans le soutien des autres.



Vous devez penser à beaucoup de plus de choses, et la qualité seule de votre site ne peut pas suffire, il va aussi falloir s'impliquer ailleurs en vous faisant connaître sur les réseaux sociaux, et en incitant les autres Webmasters à faire des liens vers votre site web. Vous aurez également l'occasion de faire profiter les autres Webmasters du succès de votre site, nous verrons tout cela dans la troisième partie de ce cours. Tout cela pour dire que le référencement dans le Web 2.0 est aussi un référencement communautaire. Votre propre qualité est importante, mais sans les autres vous n'irez pas loin.



Le référencement du Web 3.0

Se référencer sur le Web 3.0 sera aussi très différent de ce à quoi on était habitué sur le Web 2.0. En effet, vous savez d'ores et déjà que le Web 3.0 sera un web de données. On ne parle plus d'information au sein d'une page web, mais plutôt de données réelles et cohérentes. Pour se référencer correctement sur le web sémantique, il faudra donc trouver le moyen de définir des données dans vos pages web.

Nous utiliserons pour cela les Microformats dont j'ai parlé plus tôt. Comme beaucoup je pense que nous sommes à un stade de transition entre le web 2.0 et le web 3.0, c'est pourquoi il va falloir commencer sérieusement à anticiper le futur et commencer à définir des données dans vos pages web. Le Site du Zéro le fait déjà, et Google exploite déjà cela dans ses résultats de recherche... si ! Je suis sûr que vous avez déjà vu un résultat de recherche qui diffère sensiblement du modèle que nous avons vu plus tôt dans le cours.

Apprenez à programmer en Java - Java - Site du Zéro

[www.siteduzero.com/tutorial-3-10601-apprenez-a-programmer-en-ja...](http://www.siteduzero.com/tutorial-3-10601-apprenez-a-programmer-en-java...)

★★★★★ Note : 3.5/4 - 1098 votes

← Une donnée !

Le Site du Zéro met

23 juil. 2012 – Cours complet en quatre parties : bases, orienté-objets, événementiel et motifs de conception.

en forme ses résultats de recherche grâce aux Microformats

En effet, vous pouvez voir qu'une donnée est définie dans ce résultat de recherche. Les notes sur le tutoriel de Java ont été repérées par Google grâce à un Microformat : le moteur de recherche a donc été capable d'afficher des étoiles, la note /4 ainsi que le nombre de votants. Bien sûr, dans quelques années, quand le web sémantique sera parfaitement adopté, tout cela ira bien plus loin et s'étendra bien au delà des moteurs de recherche. En attendant, il y a bien d'autres données que nous pouvons définir sur notre site, Google peut s'en servir afin de mettre en forme nos résultats de recherche. Nous pouvons par exemple afficher un fil d'Ariane à la place de l'URL de la page, le prix d'un produit que nous vendons dans le cas d'un site d'e-commerce, ou encore la photo et le nom de l'auteur d'un article !

C++ - Site du Zéro

[www.siteduzero.com > Cours > Tutoriels > Programmation > C++](http://www.siteduzero.com/Cours/Tutoriels/Programmation/C++)

Tutoriels concernant la programmation en C++. ... Le Site du Zéro - Informatique. >

Cours. > Tutoriels. > Programmation. > C++. > Liste des tutoriels ...

Fil d'Ariane ajouté grâce aux micro-données !

Nous pouvons ajouter un Fil d'Ariane à la place de l'URL grâce aux micro-données !

Pour le moment, tout cela va nous servir à inciter l'internaute à visiter notre site une fois qu'il verra notre lien sur Google étant donné que nous avons l'occasion ici d'afficher plus d'informations que d'habitude.

HTML5 et les micro-données de l'avenir

Microformats ou micro-données ?

Utilisons les micro-données

Nous savons comment le Web 3.0 fonctionne et fonctionnera, pour mettre notre site aux normes il va falloir utiliser les microformats qui sont donc un balisage spécial de notre contenu HTML.

Il faut néanmoins faire la différence entre « microformats » et « micro-données ». Ces dernières ont été apportées par HTML5, et il convient de les utiliser dès maintenant et de laisser tomber les anciens microformats. Les micro-données sont amenées à devenir les références dans le futur, ce sont un peu les dernières versions en date qu'il faut vraiment utiliser : ce chapitre sera donc essentiellement porté sur les micro-données de HTML5.

Balisez sémantiquement votre contenu

Qu'est-ce qu'une micro-donnée ?

Concrètement, une micro-donnée n'est rien d'autre qu'un élément HTML un peu spécial. En fait, c'est quelque chose qui va venir se greffer sur les balises HTML déjà existantes pour les définir comme données. Ainsi, n'importe quelle balise dans votre page web peut devenir une donnée si vous souhaitez la définir comme telle, il suffit juste d'y appliquer un attribut spécifique, et le tour est joué. Étant donné que les balises `<div>` et `` sont des balises joker qui n'agissent pas sur le rendu de la page dans le navigateur, on a pour habitude de les utiliser pour définir des micro-données.

Pour définir une donnée, il suffit d'ajouter l'attribut `itemscope` à une balise. L'attribut `itemscope` sert à indiquer aux robots que le contenu de la balise qui porte cet attribut aura des données sémantiques. Elle sert donc à dire : « ok, les données qu'il y a dedans sont sémantiques, je vais pouvoir les étudier pour les prendre en compte au mieux ».

Dans la suite des exemples, nous allons tenter de définir une donnée qui représente une personne nommée Jacques grâce aux micro-données. Cela serait par exemple utile dans le cadre d'un blog pour dire qui est l'auteur de ce blog.

Code : HTML

```
<div itemscope>
  <p>Je suis l'auteur de ce blog. Mon nom est Jacques.</p>
</div>
```

Définissez le type de vos données

Vous savez maintenant définir une donnée, le problème est que votre donnée ne correspond encore à rien. Il faut dire que la `<div>` contient une personne. Pour cela on va passer par un autre attribut : `itemtype`. Cet attribut va prendre la définition d'une donnée (on parle de « vocabulaires ») depuis le site schema.org, vous pouvez consulter la liste des définitions disponibles dans la [documentation](#). Schema.org est un site maintenu par Google, Bing! et Yahoo!.

Code : HTML

```
<!-- Les moteurs de recherche sauront maintenant que cette donnée
est une personne .-->
<div itemscope itemtype="http://schema.org/Person">
  <p>Je suis l'auteur de ce blog. Mon nom est Jacques.</p>
</div>
```

Utilisez les propriétés des données !

Plusieurs propriétés sont disponibles, vous pouvez en consulter la liste exhaustive pour chaque définition dans la [documentation](#). Pour les personnes, nous allons en avoir plusieurs qui vont être intéressantes ([Consultez la documentation de la définition "Person"](#)).

- **name** :
Pour renseigner le nom de la personne.
- **addressLocality** :
Pour renseigner l'adresse de la personne.
- **affiliation** :
Pour dire si la personne fait partie d'une école, ou d'une équipe.
- **email** :
L'email de la personne.
- ...

Bref! Il y en a vraiment des tonnes et vous devriez consulter les documentations pour vous faire une idée plus complète sur la question... mais pour le moment, nous devons finir de baliser notre code. Pour cela, nous allons faire simple et utiliser le nom et l'email, cela sera bien suffisant pour compléter la présentation de Jacques sur son blog. Pour définir les propriétés, je vous conseille d'utiliser des balises ``. Utilisez l'attribut `itemprop` pour renseigner les attributs.

Code : HTML

```
<div itemscope itemtype="http://schema.org/Person">
  <p>Je suis l'auteur de ce blog. Mon nom est <span
itemprop="name">Jacques</span>.</p>
  <p>Mon adresse email est : <span
itemprop="email">jacques@monblog.com</span></p>
</div>
```

Et c'est tout ! Voilà ce que donne un code balisé sémantiquement ! Bien sûr, l'intérêt aujourd'hui est très limité et vous ne devez sûrement pas comprendre l'utilité que tout ceci peut avoir : c'est simplement de l'anticipation, car demain on ne fera plus que ça, vous devriez donc prendre l'habitude tout de suite d'utiliser les micro-données dans vos codes source. Retenez simplement que les microformats constituent l'avenir du web, et que l'on pourra plus tard utiliser ça de manière bien plus poussée qu'en 2012 : le tout est de prendre l'habitude de le faire dès maintenant ! Si vous avez compris ce concept, alors vous avez tout compris : pour le reste, je vous conseille d'aller lire la documentation de Schema.org, ainsi que le [FAQ de Google pour Webmasters à ce sujet](#).



Notez que les balises structurantes de HTML5 dont on a déjà parlé représenteront un réel enjeu d'un point de vue sémantique. Je vous conseille à nouveau de les utiliser dès maintenant pour structurer vos pages web !

En résumé:

- La sémantique représente l'avenir du web.
- Consultez les définitions sur schema.org
- Prenez dès maintenant l'habitude de baliser vos contenus grâce aux micro-données.

Les pièges à éviter et les bonnes habitudes à prendre

Les techniques de référencement sont multiples et vous savez déjà que vous allez devoir batailler sur plusieurs fronts. Il y a cependant plusieurs choses qu'il ne faut surtout pas faire dans votre référencement. Nous allons tout de suite voir les mauvaises pratiques et plus largement les pièges à éviter, mais aussi les bonnes habitudes à prendre.

Nous parlerons aussi dans ce chapitre des sites web fonctionnant grâce à Flash, une technologie signée Adobe et largement différente de HTML : les sites en Flash se référencent différemment des sites en HTML, il serait donc dommage de ne pas en parler.

Les mauvaises pratiques

Le « Duplicate Content »

D'un site à l'autre !

Google met un point d'honneur à valoriser les contenus qui présentent de l'intérêt sur le web. Mais je ne vous apprend rien en vous disant qu'il est fréquent de voir des sites web qui copient le contenu d'autres sites. Google considère ce comportement comme particulièrement malveillant et cherche à tout prix à combattre le plagiat. Ainsi, si le moteur de recherche repère un contenu plagié d'un site à l'autre, il n'hésite pas à pénaliser très lourdement le site qui a commis le plagiat. Cela implique que Google doit pouvoir :

- Identifier des contenus identiques sur le web ;
- Déterminer quel est le site original, et quel est le site qui plagie ;
- Pénaliser (lourdement) le site au contenu plagié.



Vous devez néanmoins savoir que Google a vraiment du mal à repérer les contenus copiés. Il utilise un système assez basique qui se base sur les blocs dans les pages web. Si des blocs importants en terme de volume sont strictement identiques, alors Google peut y voir un contenu plagié et considérer ce comportement comme malveillant.

Le contenu dupliqué sur un même site

Cette idée est aussi valable sur votre propre site, lorsque vos pages proposent des contenus identiques, Google peut considérer cela comme du doublon et vous pénaliser. Il faut néanmoins nuancer cette règle : certains systèmes de forums prêts à l'emploi proposent une version « desktop » ainsi qu'une version mobile indépendante. Certains Webmasters ont peur d'utiliser ces forums de peur que Google y voit un contenu dupliqué. Dans ce cas-là ne vous inquiétez pas, car Google ne considère pas ce comportement comme malveillant. Si l'on fait abstraction de ces cas spéciaux, veillez quand même à ne jamais proposer de contenus dupliqués sur votre site, c'est généralement très mauvais pour le référencement d'un site web.

Eviter d'être pénalisé : le rel="canonical"

Malgré tout, il se peut que vous n'avez pas le choix et que votre site propose des pages aux contenus identiques. Il existe un moyen de ne pas encaisser de pénalité de la part de Google, il suffit de dire au Googlebot que telle ou telle page est votre version favorite, et que vous aimeriez que Google considère celle-ci plutôt que l'autre.

Nous identifions donc :

- Une page A considérée comme page principale que Googlebot doit traiter en priorité.
- Une page B au contenu identique à la page A, nous souhaitons que Googlebot ne se focalise pas dessus.

Pour indiquer au Googlebot de ne pas se focaliser sur la page B, il suffit d'ajouter une balise dans la section `<head>` de votre page web.

Code : HTML

```
<!DOCTYPE html>
<!-- Code de la page B -->
<html>
<head>
  <meta charset="utf-8" />
  <link rel="canonical" href="a.html" />
```

```
<title>Je suis la page B !</title>
</head>
</html>
```

À la ligne 6, nous utilisons la balise `<link rel="canonical" />`. Cette balise sert à indiquer à Googlebot qu'il s'agit ici d'une page mineure : dans l'attribut href de cette balise, nous indiquons le lien de la page favorite que Googlebot devra considérer en priorité. Selon Matt Cutts, même si cette méthode est relativement propre, vous perdrez quand même un peu de Google Juice au passage. C'est pourquoi l'idéal est de ne vraiment jamais avoir de contenu dupliqué.

C'est tout ce que vous deviez savoir sur les « Duplicate Content », et sinon, n'oubliez pas que rien n'est plus efficace qu'un bon `<meta name="robots" content="noindex" />`. 🤖

Votre contenu est votre force

Interdisez l'indexation des pages pauvres en contenu

Le « Duplicate Content » n'est pas la seule erreur à ne pas faire en SEO. Si Google aime énormément vos pages qui proposent un contenu pertinent, il déteste les pages qui ne présentent aucun contenu. Prenons l'exemple des pages de contact : chaque site web en possède une en principe, il s'agit généralement d'un formulaire HTML tout simple pour envoyer un mail au webmaster.

C'est très utile pour vous car cela vous permet d'avoir des retours de vos visiteurs, seulement Google a une véritable haine contre ces pages qui pour lui ne représentent absolument rien. Google aime votre contenu textuel, et pour lui, un formulaire HTML est tout sauf un contenu textuel. Il va donc falloir dès maintenant songer à interdire l'indexation de vos pages de contact comme vous savez le faire. Voici un exemple de pages dont vous avez tout intérêt à interdire l'indexation pour ne pas recevoir de pénalités :

- Les pages de contact
- Les profils des membres

Flash et le référencement

Flash et le référencement

Qu'est-ce que Flash ?

Flash est une technologie créée par Adobe, une entreprise américaine éditant des logiciels assez chers mais de grande qualité. Ils sont très utilisés dans le monde professionnel, on peut citer Photoshop, Dreamweaver ou encore AfterEffect. Ce sont tous des produits signés Adobe dont vous avez au moins entendu le nom ! Parmi ces logiciels, il en existe un autre baptisé « Adobe Flash ». Ce logiciel et le langage qu'il propose, l'ActionScript, permettent de créer une multitude de choses complexes. À l'aide de Flash, vous pouvez notamment créer :

- Des lecteurs audio
- Des lecteurs vidéo
- Des bannières animées
- Des applications complètes
- Des jeux-vidéo



Si Flash est suffisamment puissant pour créer toutes ces choses, il permet aussi de créer des sites web complets et fonctionnels orientés contenus.

Flash vs HTML

Si vous optez pour Flash (pourquoi pas ?) il va bien falloir vous débrouiller pour référencer vos sites, car Flash est vraiment différent de HTML. Le web se fait en HTML, c'est un standard que le W3C « impose », de même que le CSS pour la mise en forme. Ainsi, tout le monde est censé utiliser HTML et CSS pour faire des sites web. Google le sait très bien, et Googlebot a été créé de telle sorte qu'il soit parfaitement adapté à l'indexation des contenus HTML : tout ça, vous le savez déjà.

Le fait est que l'on peut aussi utiliser Flash pour faire un site web, mais avec Flash, il n'y a pas de code source, le contenu n'est

pas balisé. Faites l'essai et naviguez sur [cette application Flash](#) complète. On voit bien que cela pourrait faire office de site web orienté contenu, c'est divisé en sections et il y a beaucoup de contenus textuels, cependant essayez d'afficher le code source HTML de ces pages en appuyant sur Ctrl + U... rien ne s'affiche, car comme je vous l'ai dit Flash ne fonctionne pas comme ça ! La question que vous devez donc vous poser, c'est comment référencer un site qui n'est pas composé de HTML.

La collaboration Google/Yahoo!/Adobe

Google et Yahoo! ont compris qu'ils ne pouvaient pas se permettre de ne pas savoir indexer du contenu en Flash tant ils sont nombreux sur le web. Cela a pourtant été longtemps le cas, à tel point que Flash s'est constituée une très mauvaise réputation lorsqu'il s'agissait de parler référencement : on préférerait tout simplement passer Flash sous silence et ne penser qu'en HTML et CSS. Aujourd'hui ce n'est plus tellement le cas, et officiellement, on arrive maintenant à référencer du contenu Flash pour peu que celui-ci soit bien conçu. En effet, Google, Yahoo! et Adobe collaborent depuis 2008 afin d'optimiser Flash pour le rendre « SEO-Friendly ». [Google a d'ailleurs annoncé avoir implémenté Flash](#) : il serait donc maintenant capable de l'indexer au mieux. Même si les sites Flash ne proposent pas de contenu HTML, les robots d'indexation sont aujourd'hui capables d'identifier le contenu au sein même des animations Flash ! La présence d'un contenu HTML ne serait donc pas indispensable dans l'indexation.

Je suis pour ma part incapable de vous expliquer comment construire des sites « SEO-Friendly » en Flash précisément, car j'ai toujours utilisé les standards HTML et CSS. Néanmoins, le procédé reposerait sur l'emploi de certaines fonctions en ActionScript pour déterminer quel contenu correspond à quoi, à l'aide notamment d'URL et d'ancres. Au lieu de vous faire une bête traduction des informations officielles fournies par Adobe, je vous redirige directement vers [leur FAQ qui semble complète à ce sujet](#), et surtout, vers ce [tutoriel de référencement en Flash](#).

Faut-il utiliser Flash ?

Il faut quand même savoir que Flash reste plus lourd que HTML/CSS et JavaScript. Connaissez-vous les Google Doodle ? Il s'agit de logo animés que Google propose souvent sur son moteur de recherche à la place du logo habituel. Ces logos animés sont entièrement composés en JavaScript à l'aide du framework jQuery. Google doit répondre à plus d'un milliard de requêtes chaque jour, dites-vous que cela n'est pas un hasard si la firme se refuse toujours à les proposer en Flash.

Après m'être documenté sur le sujet, mon ressenti est que malgré tous les efforts d'Adobe, de Google de Yahoo!, j'ai du mal à croire que le Flash se référence aussi bien que le HTML. Je pense néanmoins que contrairement au passé, il ne faut pas tomber dans la bête diabolisation du Flash, il semble lui aussi posséder plus d'un atout dans sa manche. Pourquoi ne pas [apprendre Flash](#) ?

En résumé

- Interdisez l'indexation de vos pages faibles en contenu pour favoriser vos pages riches en contenu.
- Méfiez-vous des « Duplicate Content », ne plagiez pas les autres sites et ne proposez pas de contenu dupliqué sur le votre !
- Attention à ne pas rendre votre site dépendant de JavaScript : Googlebot ne verra pas votre contenu si son affichage se fait grâce à JavaScript en AJAX.
- Le Flash semble se référencer correctement de nos jours. Je vous encourage à lire la FAQ d'Adobe à ce sujet.

Partie 3 : Annexes et autres ressources

Dans cette partie, nous allons aborder **plus en profondeur** les technologies intéressantes que nous avons pu évoquer dans le reste du cours, mais qui ne présentaient pas suffisamment d'intérêt pour s'inscrire comme chapitre à part entière dans le cheminement du tutoriel. Vous pouvez lire ces annexes n'importe quand pour compléter votre base de connaissances sur le sujet. Il n'est pas demandé d'avoir lu le tutoriel pour les comprendre.

Découvrez Lynx Browser !

Nos amis robots tels que Googlebot travaillent d'arrache-pied pour remonter le plus d'informations possible sur vos sites web, afin de permettre au moteur de recherche de fournir une indexation la plus pertinente qui soit.

Ces robots, extrêmement nombreux à parcourir le web tous les jours, voient la toile d'une autre manière que nous, humains. Nous allons découvrir dans cette annexe un programme nommé « Lynx Browser » qui permettra à tout le monde, pendant quelques instants, de se plonger en immersion totale [ou presque] dans la vie passionnante d'un robot. 😊

Lynx : un navigateur !

Pour ceux qui n'ont pas trop idée de ce que peut être le Lynx Browser, je préfère le leur dire tout de suite : Lynx Browser est un navigateur web, au même titre que des programmes tels que Mozilla Firefox ou Google Chrome ! 😊

Le lecteur d'écran Lynx

Petite histoire du navigateur Lynx

Lynx a été créé en 1992 pour les systèmes UNIX, mais sa popularité a fait qu'il a pu être porté pour les systèmes MS-DOS puis sur la famille des Windows, jusqu'à Windows XP. Des versions pour Windows Vista et Windows Seven ne semblent malheureusement pas exister pour le moment. Il est aujourd'hui distribué gratuitement sous licence GNU GPL et est supporté par un groupe de développeurs travaillant bénévolement sur le projet. Le rôle de cette annexe va être de vous apprendre à utiliser Lynx pour naviguer sur vos pages web d'une nouvelle manière ; vous allez voir, cela va être... très intéressant.



Je n'ai pas très envie de changer moi... Je veux bien croire que Lynx est un navigateur intéressant, mais je suis déjà très satisfait de Firefox.

Ne vous en faites, je n'allais pas vous demander de changer de navigateur web du jour au lendemain, et pour tout vous dire, Lynx Browser n'est pas vraiment prévu pour être utilisé tous les jours, il n'est pas super d'un point de vue **expérience utilisateur**.

Utilisez Lynx !

Lynx Browser est un navigateur sans interface visuelle. Si vous utilisez beaucoup Linux, vous utilisez très probablement sa bien connue console, et bien Lynx Browser **ressemble très fortement à une console Linux**. Et pour cause, sous Linux on l'utilise uniquement en mode console. Voici par exemple une capture d'écran du Site du Zéro visité sous **Lynx Browser**, réalisé sous un système Windows XP.

```

http://www.siteduzero.com/Templates/xml/news_fr.xml
# Le Site du Zéro, site communautaire de tutoriels gratuits pour débu.. (p1 of 13)
#News du Site du Zéro Derniers tutoriels du Site du Zéro index Accueil Site Du Zero
Rechercher Accessibilité

Aller au menu - Aller au contenu

* 2 970 Zéros connectés
* 288 685 Zéros inscrits
* Inscription
* Connexion
  + Pseudo
  + _____
  + Mot de passe
  + _____
  + Mot de passe oublié ?
  + [ ] Connexion auto. Connexion
  +
  + Connexion avec un compte :
    o Connexion avec compte Google
    o Connexion avec compte Facebook
  +
* Informatique
  + Informatique
  + Sciences
*
  + Cours
  + Forums
  + Participez
  + Etudes
  + Emploi
  + Boutique
* Rechercher _____ [Cours____] Go +

Les cours phares
Site Web
* HTML5 / CSS3 Ce cours existe en livre
* Javascript
* PHP débutant Ce cours existe en livre
* PHP objet

Programmation
* Langage C Ce cours existe en livre
* Langage C++ Ce cours existe en livre
(NORMAL LINK) Use right-arrow or <return> to activate.

```

Le Site du Zéro affiché par Lynx Browser



Comme vous vous en doutez, la souris ne vous servira pas à grand-chose ici, tout va se gérer au clavier !

Vous pouvez voir que Lynx Browser est un navigateur qui va récupérer des pages web et vous les afficher comme si celles-ci n'étaient reliées à aucune feuille de style, ne possédaient aucun JavaScript et ne possédaient non plus aucune image. Exactement comme Googlebot !



Pour être exact, Lynx fait partie d'une famille de logiciels que l'on appelle des « lecteurs d'écran ».

J'en parle volontairement dans ce cours, car Lynx s'avère notamment très utile pour pouvoir consulter un site internet à tout moment de la même manière que le ferait un Googlebot sans aucune limitation. Vous avez donc un programme très bien fait, robuste et léger disponible à tout moment sur votre machine.

Télécharger et installer Lynx Browser

Lynx Browser est souvent difficile à installer pour les débutants, notamment parce qu'il requiert une compilation, qui est une manipulation complexe si vous n'avez pas de bases en programmation. En outre, le lecteur d'écran est uniquement disponible sous les anciennes versions de Windows, Windows Vista et Windows Seven ne sont pour le moment pas supportés.



Heureusement, le site *Csant.info* met à la disposition de tous un installateur compilé pour les utilisateurs de Windows 9x/NT/2000/XP, à [cette adresse](#). Pour les utilisateurs de Windows Seven, le mode de compatibilité XP devrait suffire. Une fois le programme installé sur votre machine, lancez-le. Vous devriez arriver à l'écran d'accueil de Lynx Browser.

Voyons maintenant comment l'on va pouvoir visiter notre première page web sous Lynx.

Votre première visite sous Lynx Browser !

Nous allons donc voir dans cette partie du cours comment nous allons pouvoir visiter notre première page, et surtout naviguer dessus à l'aide de Lynx Browser. Vous imaginez qu'utiliser un tel navigateur diffère un peu de ce que vous avez l'habitude de faire.

La navigation sous Lynx Browser

Atteindre sa première page avec Lynx Browser !

Nous allons donc ouvrir le navigateur et nous rendre sur le **Site du Zéro**. Pour ce faire, ouvrez le navigateur. Jetez un coup d'œil au bas de la page d'accueil qui s'est ouverte devant vous. Vous devriez voir ce message.

Code : Console

```
Commands: Use arrow keys to move. '?' for help. 'q' to quit. '<-  
' to go back.
```

On vous dit ici d'utiliser les flèches de votre clavier pour vous déplacer, de taper sur la touche Q si vous voulez quitter le navigateur et de taper <- pour vous rendre à la page précédente. Je vous conseille de taper sur la touche ? afin d'ouvrir la page d'aide qui fait également office de documentation, nous ne la détaillerons pas ensemble mais il est utile de savoir qu'elle existe.

Vous pouvez aussi taper sur la touche G pour avoir la possibilité de taper l'adresse du site web sur lequel vous souhaitez vous rendre. C'est ce que l'on va faire ici. tapez donc sur la touche G. Le message que vous allez voir sera celui-ci.

Code : Console

```
URL to open:
```

On vous demande ici l'URL que vous souhaitez ouvrir dans Lynx Browser. Tapez donc "www.siteduzero.com".



Par défaut, Lynx Browser utilise le protocole HTTP. Vous n'avez donc pas à renseigner [http://](#) au début des adresses que vous souhaitez ouvrir.

Le système est sensible à la casse, vous allez donc devoir taper exactement l'URL du site que vous souhaitez visiter, si vous vous trompez, Lynx n'est pas capable de vous fournir une liste de site dont l'adresse ressemble à ce que vous avez tapé. Faites donc bien attention à ce que vous écrivez. Les prochains messages vous demanderont si vous souhaitez accepter les cookies que le Site du Zéro voudra écrire sur votre ordinateur. Donnez votre accord en tapant sur Y.



Un cookie n'est pas nocif pour votre machine.

À ce moment-là, vous allez arriver sur la page sur laquelle nous étions tout à l'heure, la page d'accueil du Site du Zéro telle qu'elle

sera vue par le Googlebot !

The screenshot shows a Lynx browser window with the address bar displaying `http://www.siteduzero.com/Templates/xml/news_fr.xml`. The main content area shows the website's navigation menu and statistics. The text is rendered in a monospaced font with green and red colors for emphasis. The menu items are listed with asterisks and plus signs, and there are search and navigation options at the bottom.

```

# Le Site du Zéro, site communautaire de tutoriels gratuits pour debu.. (pi of 13)
#News du Site du Zéro Derniers tutoriels du Site du Zéro index Accueil Site Du Zero
Rechercher Accessibilité

Aller au menu - Aller au contenu

* 2 970 Zéros connectés
* 288 685 Zéros inscrits
* Inscription
* Connexion
  + Pseudo
  + _____
  + Mot de passe
  + _____
  + Mot de passe oublié ?
  + [ ] Connexion auto. Connexion
  +
  + Connexion avec un compte :
    o Connexion avec compte Google
    o Connexion avec compte Facebook
  +
* Informatique
  + Informatique
  + Sciences
+
+ Cours
+ Forums
+ Participez
+ Etudes
+ Emploi
+ Boutique
* Rechercher _____ [Cours____] Go +

Les cours phares
Site Web
* HTML5 / CSS3 Ce cours existe en livre
* Javascript
* PHP débutant Ce cours existe en livre
* PHP objet

Programmation
* Langage C Ce cours existe en livre
* Langage C++ Ce cours existe en livre
(NORMAL LINK) Use right-arrow or <return> to activate.
  
```

Comme vous le savez, vous ne pouvez pas utiliser votre souris sur Lynx Browser. Tout ce que vous pouvez utiliser pour interagir, c'est le clavier. 😊



Mettez-y un peu de bonne volonté et dites-vous que dans les années 80 les gens qui utilisaient des ordinateurs utilisaient toujours leur clavier. Les belles interfaces avec fond d'écran animé et effets 3D au passage de la souris, c'est quelque chose de très récent à l'échelle de l'informatique. 🤖

Reconnaître et suivre les liens !

Pour naviguer entre les différents liens présents sur la page, on va devoir se servir des flèches de notre clavier. En effet grâce aux flèches Haut et Bas, vous allez pouvoir déplacer votre curseur sur toutes les zones qui sont cliquables sur les autres navigateurs. Pour « cliquer » sur un lien, tapez simplement sur la touche Entrée de votre clavier, et le navigateur vous posera les mêmes questions que tout à l'heure. (enregistrer des cookies, etc.)

Les formulaires sous Lynx Browser

Lynx va également vous permettre de remplir des formulaires. Par exemple, sur la page de connexion, vous allez pouvoir vous

positionner sur les champs `<input />` et de là vous pourrez taper directement vos informations de connexion puis valider comme vous l'auriez fait sur la version visuelle du site.

```

Password
# Connexion (p9 of 10)
> Le Site du Zéro - Informatique > Rejoignez-nous
Rejoignez-nous
Vous avez aimé les cours du Site du Zéro ?
Vous voulez écrire un tutoriel, une news ou participer aux forums ? Pour cela rien de plus simple !
Connectez-vous ou inscrivez-vous en quelques clics et rejoignez la communauté !
Connexion
= Obligatoire
Pseudo Nassoub
Mot de passe *****
Connexion automatique [ ]
Se connecter
J'ai oublié mon mot de passe !
J'ai à nouveau besoin du mail de validation !
Ou
Connectez-vous à partir de :
* Connexion avec compte Google
* Connexion avec compte Facebook
Ou
Inscription sur le Site du Zéro
Pseudo _____
Mot de passe _____
Retapez le mot de passe _____
Adresse email _____
Cette adresse sera utilisée pour valider votre compte.
Elle ne sera pas diffusée, et utilisée uniquement sur ce site.
Autoriser les partenaires du site à me contacter [ ]
Ce que nous faisons : vous inviter à des concours de programmation, vous proposer des offres d'emploi dans le développement web, etc.
Ce que nous ne faisons pas : envoyer des pubs de voyance de Mme Irma, des incitations à agrandir vos organes...
Vous pourrez vous désinscrire à tout moment depuis votre compte membre.
>Password entry field) Enter text. Use UP or DOWN arrows or tab to move off.

```

La page de connexion au Site du Zéro vue sous Lynx Browser !

L'idée est ici de proposer un navigateur qui peut remplir toutes les tâches usuelles que les autres navigateurs peuvent remplir, mais en mode pur texte. Vous vous rendez un peu mieux compte de ce que fait Googlebot maintenant. Grâce à ce navigateur, vous allez pouvoir vraiment peaufiner ce qu'il se passe sur les pages que va voir le Googlebot. Cela vous aidera par exemple à mieux choisir la position dans laquelle vous souhaitez placer vos mots-clés !

Vous pouvez bien sûr quitter à tout moment le logiciel en tapant sur la touche Q.



Le site delorie.com propose un lecteur d'écran similaire à Lynx Browser qui s'exécute directement sur votre navigateur. Il est néanmoins sujet à des limitations et a quelques restrictions et est donc moins sympathique à utiliser que Lynx Browser.

En résumé

- Lynx Browser est un lecteur d'écran. Ce sont des programmes notamment très utilisés par les internautes non-voyants.
- Lynx Browser vous permet de naviguer sur le web de la même manière que Googlebot de manière illimitée.

Les mises à jour de l'algorithme de Google étant relativement fréquentes (parfois plus de 10 mises à jour par mois) la fin de ce tutoriel ne signifie pas la fin de votre apprentissage. Les choses bougent : vous devrez donc constamment vous tenir prêt à apprendre de nouvelles choses. J'espère en tout cas que vous avez maintenant de quoi bien commencer dans le monde du référencement. Qui sait, peut-être aurais-je le plaisir de retrouver votre site dans le Triangle d'Or ?

Ce tutoriel est en phase de rédaction. De nombreuses mises à jour sont à venir.