

DESIGN  
WEB  
UX

Préface d'Olivier **Andrieu**

Alexandra **Martin**  
Mathieu **Chartier**

Techniques de  
**référencement**  
web

2<sup>e</sup> éd.

AUDIT ET SUIVI SEO

EYROLLES

# Techniques de référencement web 2<sup>e</sup> éd.

« Ce duo de choc et de charme vous livre les clés techniques du référencement dont vous devrez tenir compte pour un meilleur positionnement.

Un livre incontournable ! »

Isabelle Canivet

Experte en stratégie de contenu et SEO  
<http://www.yellowdolphins.com>

« Le monde SEO francophone manquait d'un ouvrage technique destiné aux développeurs n'étant pas obligatoirement spécialistes des arcanes du référencement naturel. »

Olivier Andrieu

Expert en référencement  
<http://www.abondance.com>

## Un livre technique pour les référenceurs

L'objectif de cet ouvrage est de répondre aux nouvelles contraintes imposées par le monde du référencement web. Non pas que le **SEO** change radicalement de manière quotidienne, il faut bien avouer que des nouveautés ne cessent d'être annoncées et nous devons être de plus en plus aguerris pour réagir face à ces évolutions.

Après quelques sains rappels sur l'historique des moteurs de recherche et la méthodologie du référencement, ce livre se positionne en effet de façon beaucoup plus technique que les autres sur le sujet. On y découvre par exemple les techniques à maîtriser pour créer un Sitemap ou un fichier robots.txt : c'est un véritable **manuel de programmation** pour référenceur. S'ajoutent à cela deux parties indispensables pour gérer l'audit et le suivi SEO d'un site web.

Un **ouvrage essentiel** pour les chefs de projet web, référenceurs et développeurs ! Enrichie de nombreux exemples et de nouvelles sections sur le référencement d'applications et de sites mobiles, la sécurité et le SEO local, la seconde édition de cet ouvrage est également 100 % à jour sur les derniers algorithmes de Google !

Autodidacte chevronnée, Alexandra Martin, aka Miss SEO Girl, est l'auteur du blog [www.miss-seo-girl.com](http://www.miss-seo-girl.com). D'un profil webmarketing et spécialiste du référencement naturel, ses compétences sont complétées par celles de Mathieu Chartier, véritable passionné par la technique et le référencement dès ses premières heures, et auteur de plusieurs livres et du site [blog.internet-formation.fr](http://blog.internet-formation.fr).

## AU SOMMAIRE

**Historique de Google et Bing.** La programmation au service du référencement  
⊗ **Maîtriser les techniques d'indexation.** Rappel des fondamentaux  
⊗ Maîtriser les Sitemaps XML  
⊗ App Indexing  
⊗ Créer un fichier robots.txt  
⊗ **Optimiser le positionnement par la technique.** Gérer les fichiers .htaccess et les serveurs Apache  
⊗ ASP, ASP.Net et configuration des serveurs IIS de Microsoft  
⊗ Compatibilité mobile  
⊗ Sécurité avec HTTPS  
⊗ AMP HTML  
⊗ Créer un système de hashtags optimisé SEO avec PHP  
⊗ AuthorShip et AuthorRank  
⊗ SEO local  
⊗ **Facteurs bloquants et pénalités Google.** **Le suivi du référencement.** Google Analytics et ses secrets  
⊗ Analyse qualitative et ROI  
⊗ **L'audit SEO.** L'audit technique  
⊗ Check-list de l'audit SEO

Techniques de  
**référencement**  
**web**

AUDIT ET SUIVI SEO

## DANS LA MÊME COLLECTION

C. LALLEMAND, G. GRONIER. – **Méthodes de design UX.**

N°14143, 2015, 488 pages.

S. DAUMAL. – **Design d'expérience utilisateur.**

N°14176, 2<sup>e</sup> édition, 2015, 220 pages.

S. POLLET-VILLARD. – **Créer un seul site pour toutes les plates-formes.**

N°13986, 2014, 144 pages.

K. DELOUMEAU-PRIGENT. – **CSS maintenables avec Sass et Compass.**

N°13640, 2<sup>e</sup> édition, 2014, 252 pages.

J. PATONNIER, R. RIGOT. – **Projet responsive web design.**

N°13713, 2013, 162 pages.

I. CANIVET, J.-M. HARDY. – **La stratégie de contenu en pratique.**

N°13510, 2012, 176 pages.

C. SCHILLINGER. – **Intégration web – Les bonnes pratiques.**

N°13370, 2012, 390 pages.

## SUR LE MÊME THÈME

O. ANDRIEU. – **Réussir son référencement web.**

N°14118, édition 2016 (à paraître).

O. ANDRIEU. – **SEO zéro euro.**

N°14033, 2014, 224 pages.

I. CANIVET. – **Référencement mobile.**

N°13667, 2013, 456 pages.

I. CANIVET. – **Bien rédiger pour le Web.**

N°13750, 3<sup>e</sup> édition, 2014, 736 pages.

D. ROCH. – **Optimiser son référencement Wordpress.**

N°14182, 2<sup>e</sup> édition, 2015, 296 pages.

E. MARCOTTE. – **Responsive web design.**

N°13331, 2011, 160 pages.

F. DRAILLARD. – **Premiers pas en CSS 3 et HTML 5.**

N°13944, 6<sup>e</sup> édition, 2015, 472 pages.

Retrouvez nos bundles (livres papier + e-book) et livres numériques sur

<http://izibook.eyrolles.com>

Alexandra **Martin**  
Mathieu **Chartier**

# Techniques de référencement web

2<sup>e</sup> éd.

AUDIT ET SUIVI SEO

Préface d'Olivier **Andrieu**

EYROLLES

Éditions Eyrolles  
61, bd Saint-Germain  
75240 Paris Cedex 05

En application de la loi du 11 mars 1957, il est interdit de reproduire intégralement ou partiellement le présent ouvrage, sur quelque support que ce soit, sans l'autorisation de l'Éditeur ou du Centre Français d'exploitation du droit de copie, 20, rue des Grands Augustins, 75006 Paris.

# Préface

---

Je me souviens d'un jour de septembre 1993. Après un long voyage en avion et un décalage horaire de nombreuses heures, je sortais des bureaux d'Eyrolles après avoir signé le contrat de mon premier ouvrage, *Internet guide de connexion*, qui allait sortir l'année suivante. Vous vous rendez compte, un livre entier sur les différentes façons de se connecter à Internet... Un sujet qui rendrait le moindre internaute perplexe aujourd'hui... C'était pourtant il y a vingt ans à peine. Pour l'anecdote, ce livre proposait en annexe, sur quelques pages, la liste exhaustive des sites web français avec leur description... Et ce n'était pas écrit petit...

Et puis j'ai créé ma société, Abondance, en 1996, pour proposer une offre de référencement sur Excite, Infoseek et autres Lycos. C'était en période pré-googlienue :-). Autre anecdote : j'ai eu mon quart d'heure de gloire « warholien » en faisant une formation au référencement (qui ne s'appelait pas encore SEO) et au fonctionnement des moteurs de recherche chez... Altavista ! Je vous parle d'un temps que les moins de vingt ans... Vous connaissez la suite...

Pourquoi je vous parle ici tel un ancien combattant ? Parce que, durant toutes ces années, j'ai eu l'occasion de participer à des dizaines, peut-être même des centaines de conférences et de formations sur ce domaine si passionnant qu'est le SEO. Souvent, à la fin de ces « prestations », les gens viennent discuter avec vous, poser des questions, demander des éclaircissements sur certains points ou par rapport à leurs besoins spécifiques. Parfois, également, des participants vous expliquent qu'ils ont beaucoup appris en lisant vos livres, en parcourant votre site, en suivant une de vos formations. Ça fait bien sûr très plaisir, non seulement parce que ça flatte l'ego (ça, on pourrait s'en passer assez rapidement, d'autant plus que le mien n'est pas fondamentalement développé), mais surtout parce que cela vous montre que le travail que vous faites n'est pas vain et qu'il aide d'autres personnes. N'est-ce pas là l'essence même de notre existence ? De nombreux e-mails très sympathiques vous indiquent également que, parfois, un de vos ouvrages a suscité des vocations. Le plaisir que cela procure est évident, même si ça vous renvoie parfois à l'état d'ancêtre...

Alexandra et Mathieu, les auteurs du livre que vous tenez entre les mains, font partie des personnes qui m'ont contacté, un jour, pour me dire qu'ils avaient commencé à « mettre les mains dans le cambouis des moteurs » à l'aide, entre autres, d'un de mes ouvrages. Ils ont fait preuve d'une gentillesse et d'une humilité qui, au-delà des qualités humaines que l'on peut ainsi discerner, sont, selon moi, un ingrédient essentiel pour devenir un bon référenceur. Bref, ce sont « des gens bien ». J'ai été d'autant plus intéressé par leur projet qu'il me semblait combler un manque dans notre domaine : un ouvrage qui s'adresse avant tout aux développeurs, avec une vision technique du référencement. Non seulement une personne qui connaît déjà plus ou moins le monde du SEO et qui veut se perfectionner au niveau du code, mais

également l'inverse, c'est-à-dire le développeur pointu dans son métier mais qui cherche à mieux comprendre le référencement naturel. Je suis certain que ce livre répondra à cette double attente.

Car le challenge est important : depuis vingt ans que je navigue dans les méandres du Web et que je fais des audits SEO, j'ai pu me rendre compte à quel point les sites web n'étaient pas toujours développés en prenant en compte les critères de pertinence de Google – ce qui nécessite parfois de reprendre à zéro un projet. Tout cela parce que les développeurs n'étaient, le plus souvent, pas sensibilisés aux arcanes des moteurs de recherche et à leur fonctionnement. De plus, ils n'étaient pas toujours conscients de tous les outils qu'il était possible d'utiliser et de créer pour optimiser et suivre un site ainsi que sa visibilité.

Je formule donc le vœu que ce livre aide à la fois les développeurs et les spécialistes SEO aux joies du codage. Et, comme il paraît que la programmation informatique est prévue dans les années qui viennent dès le plus jeune âge au programme scolaire de nos chères têtes blondes, je souhaite le meilleur à cet ouvrage : devenir dans le futur un manuel scolaire reconnu pour que les sites web que nos enfants créeront soient nativement *search engine friendly*.

Olivier Andrieu

Éditeur du site Abondance.com



# Avant-propos

---

## Pourquoi ce livre ?

Ce livre a été rédigé afin de répondre aux dernières contraintes imposées par le monde du référencement web. Non pas que la SEO change radicalement de manière quotidienne, mais il faut bien avouer que des nouveautés ne cessent d'arriver et que nous devons être de plus en plus aguerris pour répondre à ces évolutions.

Certaines thématiques sont souvent balayées au sein des articles sur la Toile voire dans des ouvrages dédiés au référencement, les auteurs ont souhaité fournir des réponses précises et détaillées sur ces sujets en passant à la fois par la technique, la programmation et l'usage avancé d'outils disponibles sur le marché. Leur volonté a aussi été portée sur l'analyse et le suivi des données ainsi que sur l'audit de sites web pour trouver les forces et faiblesses des pages web en termes de SEO, c'est pourquoi une majeure partie du livre est consacrée à ces sujets.

Ce livre a été rédigé pour donner des explications avancées sur les évolutions des moteurs de recherche mais aussi sur des critères précis et techniques de référencement. En définitive, le contenu est partagé assez équitablement entre théorie, technique, suivi et audit de sites web. Toutefois, tous les sujets ne peuvent pas être traités avec la même minutie, ce qui explique que des choix ont été effectués pour tenter d'apporter un maximum de savoir et de consistance aux lecteurs, tout en mixant pédagogie et technique.

Nombre d'ouvrages de qualité existent sur le sujet, non sans rappeler celui d'Olivier Andrieu, de Daniel Roch ou même de Mathieu Chartier, l'un des auteurs de ce livre, mais ils s'adressent généralement aux débutants en matière de référencement. Ici, la volonté des auteurs a été d'adapter leurs connaissances pour des spécialistes, des développeurs ou webmasters qui souhaitent passer un cap dans ce métier. Non pas que les débutants soient exclus et ne puissent pas comprendre les propos tenus, certaines parties imposent toutefois une bonne connaissance technique des langages et du Web, et bien que les auteurs fassent leur maximum pour clarifier les contenus, des difficultés peuvent être ressenties.

Le livre peut être lu de manière linéaire puisque les auteurs ont suivi la logique méthodologique du référencement en débutant avant tout par l'indexation des pages web puis les optimisations du positionnement pour conclure avec des méthodes de suivi et d'audit approfondies. Pour les plus spécialistes, la lecture

peut également se faire de manière décousue afin de répondre aux besoins directs puisque les sections littéraires ont aussi une certaine indépendance.

Vous pouvez retrouver l'ensemble des codes et programmes du livre à l'adresse <http://goo.gl/onrvlx> ou sur la fiche de l'ouvrage sur [www.edition-eyrolles.com](http://www.edition-eyrolles.com).

## À propos des auteurs

Consultante en référencement, Alexandra Martin accompagne les professionnels dans la mise en place de leur stratégie de visibilité globale sur le Web depuis 2008. Formée dans le monde du marketing, la SEO est devenue une passion insatiable qu'elle partage depuis plusieurs années sur son blog [www.miss-seo-girl.com](http://www.miss-seo-girl.com) au travers d'articles variés et à la portée de tous. Retrouvez-la sur son compte Twitter @Miss\_Seo\_Girl et sur son blog pour découvrir ou redécouvrir son univers.

Mathieu Chartier est un ancien archéologue reconverti au Web à la fin des années 2000 après avoir suivi un master professionnel dans ce domaine. Passionné par la technique et le référencement dès ses premières heures, il a logiquement développé ces aspects au travers de ses ouvrages, de son blog professionnel (<http://blog.internet-formation.fr>) mais aussi dans son activité de consultant, formateur et webmaster multitâches. Vous pouvez le retrouver sur Twitter (@Formation\_web) ou sur son blog, n'hésitez pas à lui poser des questions ou à lui demander conseil en matière de SEO..



*Mathieu Chartier, Olivier Andrieu et Alexandra Martin*

## Remerciements

Nous tenons à remercier toutes les personnes qui nous ont aidés et suivis lors de la rédaction de ce livre, à commencer par les éditions Eyrolles qui nous ont donné la chance de réaliser ce rêve.

Nos premières pensées vont vers nos proches, nos amis et bien sûr notre famille dont la patience, la tolérance, le soutien et l'amour ont été nos forces pour mener à bien ce projet éditorial. Nos esprits et nos cœurs se tournent vers nos moitiés respectives, Yann et Anne-Sophie, ainsi que vers d'autres personnes bien trop nombreuses pour être marquées à l'encre noire, bien que nous puissions glisser au moins quelques noms : Gaby, Denis, Ilou, Julien, Guillaume, Rodrigo, Yannis, Elena, Chrystelle, Christophe, Sandra, Cynthia, Carine, Céline et tant d'autres sans qui nous ne saurions rien...

Par ailleurs, il nous est impossible de citer tous les référenceurs, marketeurs et développeurs qui ont pu nous aider activement ou inconsciemment mais voici tout de même une liste de noms qui comptent à nos yeux et sans qui la discipline serait fade : Guillaume Degré, Laurent Bourrelly, XavFun, Thomas Cubel, Hasni Khabeb, Ronan Chardonneau, Edouard Ouvrard, Daniel Roch, Dejan Markovic, Isabelle Canivet, Ferréole Lespinasse, Marie Pourreyron, Sandrine Khou, Sylvain Richard, Sébastien Monnier, Renaud Joly, Nicolas Robineau et tous les autres que nous n'avons plus la place de nommer...

Cette liste non exhaustive de spécialistes à qui nous tenons ne pourrait être complète sans présenter notre préfacer Olivier Andrieu, dont les livres nous ont sensibilisés au métier, et sans qui le référencement n'aurait pas la même saveur en France. Nous tenons à le féliciter pour tout ce qu'il a apporté à la sphère SEO avec une humilité sans faille et une passion communicative, mais aussi à le remercier pour son soutien et son aide.

Enfin, nous tenons à remercier l'ensemble des lecteurs de nos blogs respectifs, [www.miss-seo-girl.com](http://www.miss-seo-girl.com) et [blog.internet-formation.fr](http://blog.internet-formation.fr), et espérons que vous continuerez à nous être fidèles.

Nous vous aimons tous, merci pour votre soutien inégalable.

Alexandra Martin et Mathieu Chartier



# Table des matières

---

<b>Introduction</b> .....	1
<b>Historique de Google et Bing</b> .....	1
Quelques dates et chiffres clés côté Google/Bing .....	1
Évolutions des moteurs de recherche .....	5
Peut-on vivre sans Google ? .....	9
<b>La programmation au service du référencement</b> .....	10
Différencier les langages côté client et côté serveur .....	11
Quel rôle pour le référencement ? .....	12
Socle technique pour programmer sur le Web .....	14
Gérer les animations avec JavaScript, jQuery, Prototype et consorts.....	16
Bases de PHP .....	17
Conclusion sur la programmation .....	20
 CHAPITRE 1	
<b>Maîtriser les techniques d'indexation</b> .....	21
<b>Rappel des fondamentaux</b> .....	21
Rôle et importance de l'ergonomie .....	22
Méthodes d'indexation .....	23
Maîtriser l'évolution de l'indexation .....	24
<b>Maîtriser les Sitemaps XML</b> .....	30
Origines et usages .....	30
Étapes de création .....	31
Soumettre des fichiers Sitemap .....	31
Créer un Sitemap index .....	33
Concevoir un Sitemap XML .....	34
Autres types de fichiers Sitemap .....	35
Exemples d'outils d'aide à la création de fichiers Sitemap .....	37

Créer son propre générateur avec PHP et MySQL .....	39
<b>Créer un fichier robots.txt. ....</b>	<b>47</b>
Principe général de fonctionnement .....	47
Étapes de création d'un robots.txt. ....	48
Outils et spécificités des fichiers robots.txt. ....	52
Autres techniques d'optimisation .....	56
L'App Indexing : indexer des liens profonds d'applications mobiles .....	72
Faire du SEO local .....	75
 CHAPITRE 2	
<b>Optimiser le positionnement par la technique .....</b>	<b>85</b>
<b>Rappels des fondamentaux .....</b>	<b>85</b>
Methodologie du positionnement .....	85
Les optimisations internes .....	87
Les optimisations off page .....	94
<b>Gérer les fichiers .htaccess et les serveurs Apache .....</b>	<b>101</b>
PageSpeed et vitesse de chargement des pages. ....	103
Gérer des redirections. ....	122
Gérer les redirections spécifiques et les codes d'erreurs. ....	125
Maîtriser la réécriture d'URL. ....	129
Autres astuces avec les fichiers .htaccess .....	138
<b>ASP, ASP.Net et configuration des serveurs IIS de Microsoft. ....</b>	<b>140</b>
Faire des tests avec un serveur IIS installé localement. ....	140
Effectuer des redirections avec IIS, ASP et ASP.Net .....	142
Nettoyer les URL avec VBScript .....	143
Réécrire des URL avec un serveur Microsoft .....	144
Autres spécificités techniques du fichier web.config .....	146
<b>Compatibilité mobile .....</b>	<b>150</b>
Pourquoi posséder un site mobile-friendly ? .....	150
Différentes alternatives mobiles .....	155
Exemples de mises en application. ....	158
<b>AuthorShip et AuthorRank. ....</b>	<b>166</b>
AuthorShip .....	166
PublisherShip .....	172
AuthorRank et AgentRank .....	174
<b>Créer un système de hashtags optimisé SEO avec PHP. ....</b>	<b>179</b>

<b>Optimiser le Rank Sculpting et le Bot Herding</b> .....	183
PageRank Google .....	183
TrustRank : indice de confiance .....	185
BrowseRank de Bing .....	186
Rank Sculpting et Bot Herding .....	187
<b>Passer en HTML 5 plutôt qu'en XHTML ou HTML 4 ?</b> .....	190

## CHAPITRE 3

<b>Facteurs bloquants et pénalités Google</b> .....	193
<b>Principales mises à jour des moteurs de recherche</b> .....	193
Google Panda .....	194
Google Phantom (Quality update) .....	196
Google Penguin .....	197
Les EMD (Exact Match Domain) .....	200
Google Page Layout .....	202
Google PayDay Loan .....	204
<b>Redirections mobiles spammy pour faire de l'affiliation</b> .....	206
<b>Sites piratés</b> .....	207
<b>Qualité des contenus chez Bing</b> .....	209
<b>Facteurs bloquants et solutions alternatives</b> .....	209
Frames .....	209
Listes déroulantes avec liens HTML .....	210
Formulaires et accès limités .....	212
ActionScript et sites Full Flash .....	214
Ajax et JavaScript non optimisés .....	216
Cookies et sessions .....	220
<b>Typologie des pénalités</b> .....	223
Différencier les sanctions manuelles ou algorithmiques .....	223
Sandbox .....	224
Baisse de PageRank .....	224
Déclassement .....	225
Liste noire .....	225
Comment faire pour sortir d'une pénalité Google ? .....	226
<b>Quelques causes de pénalités</b> .....	229
Spamdexing .....	229
Keyword stuffing .....	232

Cloaking .....	233
Doorway ou pages satellites .....	236
Contenus dupliqués et DUST .....	236
Les paids links .....	242
Rich snippets abusifs .....	243
Negative SEO .....	244
CHAPITRE 4	
<b>Le suivi du référencement</b> .....	247
<b>Suivre l'indexation</b> .....	247
Voir le site avec l'œil du spider .....	248
Suivre les pages indexées .....	259
<b>Suivre le positionnement</b> .....	274
Du mouvement dans les SERP ? .....	274
Suivre les positions et les requêtes .....	278
Suivre les backlinks avec des outils .....	291
<b>Google Analytics et ses secrets</b> .....	295
Présentation et usage de l'outil .....	295
Méthodes de tracking .....	302
Filtres et rapports pour le SEO .....	310
Suivre la fréquence du crawl en direct .....	319
Peut-on contrer les not provided ? .....	319
Conclusion sur Google Analytics .....	321
<b>Analyse qualitative et ROI</b> .....	321
CHAPITRE 5	
<b>L'audit SEO</b> .....	327
<b>L'audit technique</b> .....	328
Le nom de domaine .....	328
Le fichier robots.txt .....	330
Le fichier sitemap.xml .....	332
La qualité du code source .....	332
Les URL .....	334
Compatibilité de votre site .....	335
Les erreurs 404 et leur page dédiée .....	336
Hiérarchisation et structure interne .....	338
Fil d'Ariane .....	339



Publicité et pop-ups .....	339
Logo cliquable.....	339
Favicon .....	341
Rich snippets.....	342
Hébergements et serveurs.....	342
<b>Audit de contenu .....</b>	<b>347</b>
La balise <title> .....	347
La balise meta description .....	347
L'utilisation des titres internes avec <hn> (<h1> à <h6>). .....	348
Sémantique et structure HTML .....	348
Les contenus textuels .....	351
Choix et utilisation des mots-clés.....	353
Longue traîne .....	354
Contenu dupliqué .....	354
Les contenus des médias.....	355
La fréquence de mise à jour .....	356
Le maillage interne.....	357
<b>Audit de popularité .....</b>	<b>357</b>
Analyse des backlinks .....	358
Les réseaux sociaux .....	361
<b>Techniques avancées et outils d'audit.....</b>	<b>364</b>
De bons outils sur le marché.....	364
Suivre les données avec PHP .....	369
<b>Check-list de l'audit SEO .....</b>	<b>396</b>
 ANNEXE	
<b>Sources de veille SEO .....</b>	<b>399</b>
<b>Ressources techniques.....</b>	<b>399</b>
<b>Interfaces pour les webmasters .....</b>	<b>399</b>
<b>Documentation et blogs officiels des moteurs de recherche.....</b>	<b>400</b>
<b>Antipénalités, réexamen et vie privée .....</b>	<b>400</b>
<b>Soumission manuelle aux moteurs de recherche.....</b>	<b>400</b>
<b>Sources généralistes sur le référencement.....</b>	<b>401</b>
<b>Baromètres, études chiffrées et statistiques.....</b>	<b>401</b>
<b>Simulateurs de robots d'indexation .....</b>	<b>402</b>

<b>Outils d'analyse des liens</b> .....	402
<b>Outils de recherche de mots-clés</b> .....	403
<b>Outils d'analyse des contenus et des mots-clés</b> .....	403
<b>Audit SEO, aide et suivi</b> .....	404
<b>Outils antiplagiat et duplicate content</b> .....	405
<b>Analyse du PageSpeed et de la vitesse de chargement</b> .....	405
<b>Réseaux sociaux</b> .....	406
<b>Index</b> .....	407

# Introduction

---

Tout au long de ce livre, nous allons étudier des techniques avancées en matière de référencement, parfois peu exploitées ou méconnues, afin d'être plus visible et de gagner des positions dans les moteurs de recherche.

Nous traiterons d'une multitude de sujets mais nous insisterons essentiellement sur les parties les plus techniques de la discipline ainsi que sur le suivi et l'audit d'un site web en matière de SEO. Si vous n'êtes pas encore à l'aise avec le vocabulaire relatif à ce domaine, vous pouvez vous référer à l'excellent glossaire de Jonathan Rousseau (source : <http://goo.gl/kGl0kh>) qui vous permettra d'appréhender sans aucun souci la suite de cet ouvrage.

Pour que l'entrée en matière ne soit pas trop brutale pour les moins connaisseurs d'entre vous, nous allons tout d'abord présenter dans cette introduction un rapide historique des deux moteurs principaux du marché en France, Google et Bing. Nous vous présenterons ensuite les bases techniques essentielles à maîtriser pour comprendre les codes proposés tout au long des chapitres qui composent ce livre. Nous vous souhaitons une agréable lecture...

## Historique de Google et Bing

### Quelques dates et chiffres clés côté Google/Bing

#### Historique de Google

Tout commence grâce à la passion démesurée de l'informatique de deux étudiants de l'université de Stanford, Larry Page (22 ans) et Sergey Brin (21 ans), qui se rencontrent en 1995 et deviendront rapidement des amis mais aussi de futurs grands noms du Web.

Dès 1996, ils débutent leur aventure commune avec la création d'un premier moteur de recherche nommé « BackRub ». Ce moteur fonctionnait uniquement via les serveurs de l'université de Stanford et était relativement gourmand en bande passante au point d'être abandonné à la demande de l'université.

#### Une entreprise planétaire

En 2013, la société comptait plus de 50 000 salariés à travers le monde. Tous proviennent d'horizons différents et sont polyglottes, afin de mieux représenter les utilisateurs internationaux de Google.

Les bureaux sont appelés « Googleplex » et la société en dénombre pas moins de soixante-dix dans plus de quarante pays à travers le monde, dont un à Paris.

Il ne faudra pas longtemps pour que le projet revoie le jour sous l'effigie de « Google » puisque dès le 4 septembre 1998, les deux amis créent la société éponyme à Mountain View en Californie, dans la

Silicon Valley. Les deux fondateurs engagèrent rapidement leur premier salarié, Craig Silverstein, et leur société devint la puissante entreprise que nous connaissons.

Figure I-1

*Larry Page et Sergey Brin  
dans leur jeunesse*



Depuis 2011, le PDG de Google est Larry Page. C'est la seconde fois qu'il occupe ce poste depuis la naissance de Google. Il a juste cédé sa place entre 2001 et 2011 à Eric Schmidt, ancienne tête pensante de Google et actuellement président exécutif.

Le nom Google a pour origine un terme mathématique, « googol » en anglais, désignant le chiffre 1 suivi de 100 zéros. Ce terme traduit l'ambition des deux fondateurs de gérer un volume infini d'informations sur la Toile. Une rumeur circule et précise qu'au moment de l'enregistrement, une erreur de frappe surgit et au lieu de taper « googol », le mot Google fut tapé et conservé. Le nom de domaine google.com a été déposé le 15 septembre 1997.

#### **Pourquoi tant de « O » ?**

Quand nous procédons à une recherche, Google propose un nombre incalculable de pages en bas des résultats. Dans ce cas, la lettre « O » de Google se multiplie comme le terme mathématique « googol » le désigne et devient « goooooogle ».

Le slogan de Google est « Don't be evil », traduit littéralement par « Ne soyez pas malveillants ». Il souligne la volonté de Google de toujours faire les choses correctement d'un point de vue éthique (bien que cela puisse être discuté parfois). Google tente d'appliquer cette règle au sein de sa société et demande la même chose aux référenceurs. Ne soyez pas malhonnêtes, ne trichez pas, soyez fair-play et respectez les consignes. Cette devise s'applique à la lettre et explique les nombreuses pénalités appliquées par le moteur de recherche.

En SEO, on différencie ainsi les « White Hat » (internautes « propres »), les « Grey Hat » (référenceurs qui essaient de rester dans les consignes) ainsi que les « Black Hat » (référenceurs mal intentionnés ou qui outrepassent les consignes) mais tout cela n'est en réalité qu'une histoire de jargon... Il faut surtout

retenir que Google édicte des *guidelines* qu'il est préférable de respecter, mais des Black Hat sont aussi des référenceurs qui font progresser la discipline avec leurs découvertes, il faut donc aussi admettre qu'ils jouent un rôle parfois favorable au SEO.

Google se caractérise par un historique riche, rempli d'événements majeurs. À dire vrai, la société est en perpétuelle évolution depuis sa naissance en 1998. Elle rachète sans cesse des entreprises, lance de nouveaux services ou met à jour l'existant. Nous pouvons citer notamment Analytics, AdWords, Gmail, YouTube, Chrome, Google+, etc., autant de services que Google met gratuitement à disposition de ses utilisateurs, bien que certains n'aient pas duré tels que Google Wave ou encore l'illustre agrégateur iGoogle.

Terminons notre tour d'horizon par un historique daté de Google :

- octobre 2000 : lancement de Google AdWords ;
- septembre 2002 : Google Actualités (avec déjà 4 000 sources d'actualités) est lancé ;
- avril 2004 : naissance du service Gmail ;
- août 2004 : Google annonce son entrée en bourse à Wall Street ;
- février 2005 : lancement de Google Maps ;
- juin 2005 : naissance de Google Earth ;
- novembre 2005 : mise en place de Google Analytics ;
- 2006 : lancement de Picasa (janvier), de Google Agenda (avril), de Google Trends (mai) et de Google Apps (août) ;
- octobre 2006 : rachat de YouTube ;
- novembre 2007 : lancement de l'OS mobile Android ;
- septembre 2008 : le navigateur Chrome est déployé ;
- octobre 2009 : accord avec Twitter pour insérer les tweets dans les résultats de recherche (ce partenariat n'est plus d'actualité) ;
- mars 2011 : lancement du bouton +1 pour « liker » les pages ;
- juin 2011 : naissance du réseau social Google+.

#### Plus d'informations sur l'historique

Pour en savoir plus sur l'historique détaillé de Google, vous pouvez consulter la page suivante : <https://www.google.fr/about/company/history/>.

Google est le leader du marché de la recherche web, nous le savons mais les statistiques nous le confirment constamment. Les chiffres sont éloquentes tant Google a connu une progression fulgurante dès ses premiers pas dans le monde :

- en mai 2014 en France, Google détient environ 92 % de parts de marché (source : <http://goo.gl/kZtIjH>) ;
- l'index de Google contient environ 30 trillions de documents ;
- Google met à disposition de ses usagers près de 200 produits et services (API, apps mobiles...) dont Blogger, Google Agenda, Google Earth, Google Docs, Google Alerts, etc. ;

- Google a fait l'acquisition de plus de 100 entreprises dont certaines très connues comme Picasa (juillet 2004), Keyhole (octobre 2004, devenu ensuite Google Earth), Urchin Software (mars 2005, utilisé pour créer Google Analytics), YouTube (octobre 2006), FeedBurner (mai 2007, spécialiste du flux RSS et Atom) ou encore eBook Technologies (janvier 2011).

### La stratégie de Google

L'objectif de Google est d'investir dans des entreprises spécialisées dans divers domaines d'activité afin d'étoffer son offre de produits et de services sans passer par des prestataires extérieurs. Google s'ouvre de plus en plus aux réseaux sociaux, au monde mobile, à la robotique, à la domotique et à l'e-commerce.

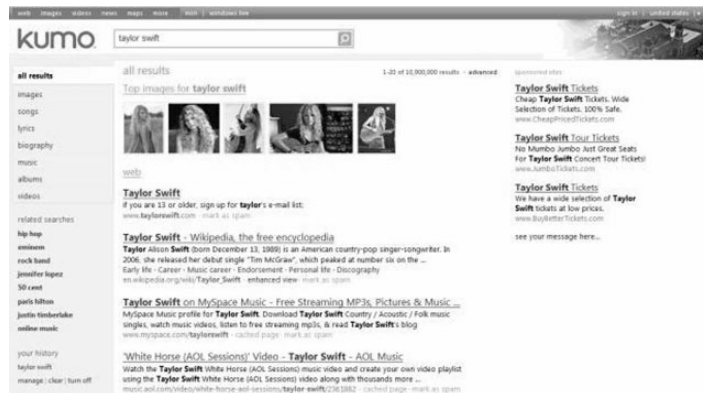
### Historique de Bing

Bing est un moteur de recherche récent développé par Microsoft après avoir mis en place des technologies de recherche telles que Live Search, Windows Live Search (abandonné dès 2011) ou encore MSN Search. Il a été lancé officiellement le 3 juin 2009.

« Bing » provient d'une onomatopée inventée par Microsoft car ce nom était simple à retenir, fonctionnel et pouvait être apparenté au son émis nativement en cas de prise de décision sur Windows. Pour l'anecdote, Bing aurait pu ne jamais exister sous ce nom puisqu'il était à l'origine surnommé Kumo dans les captures de présentation en mars 2009 (source : <http://goo.gl/zA70rd>), ce qui montre que la firme cherchait un nom court, percutant et facilement mémorisable.

Figure 1-2

*Bing aurait pu s'appeler Kumo.*



Dès le mois de juillet 2009, peu de temps après le lancement du moteur, un partenariat entre Microsoft et Yahoo! a été signé afin de fonder une alternative plus puissante et efficace contre Google. Désormais, la technologie de recherche de Bing est déployée sur le moteur de Microsoft mais aussi sur Yahoo!, bien que la gestion des liens sponsorisés soit quant à elle administrée essentiellement sur les compétences de Yahoo! et de ses ingénieurs.

En mai 2010, Microsoft a réussi à obtenir un partenariat avec les constructeurs de la marque BlackBerry afin que son moteur soit déployé de manière systématique sur ces supports mobiles.

Retenons également que le logo de Bing a été totalement revisité en 2013 à l'image de Windows 8 et de son *flat design*. Avant cette refonte, le moteur arborait un logo textuel bleu qui a été remplacé par un « b » schématique accompagné du mot « Bing » tout en jaune.

## Évolutions des moteurs de recherche

Les moteurs de recherche ont tous connu des progressions nettes ses dernières années, que ce soit Google, Bing, Yahoo!, Baidu, Ask ou Yandex. Il est impossible de détailler toutes les évolutions, mais nous allons présenter rapidement ici quelques mises à jour et services qui ont modifié le visage de la recherche web dans le monde.

### Quelques mises à jour de Google

Il est important de se rappeler que l'année 2009 a été marquante chez Google et que le moteur que nous utilisons quotidiennement n'a plus du tout la même allure... Cette année a marqué l'arrivée de la recherche en « temps réel » et de la recherche universelle avec Google Maps dans les résultats de recherche par exemple.

Parmi les évolutions les plus marquantes dans l'histoire de Google, il ne fait aucun doute que Google Caffeine, déployée dès juin 2010, a changé l'approche du moteur de recherche en termes d'indexation et de perception des pages web. Caffeine a été une mise à jour complète du système d'indexation des pages dans le moteur de recherche afin de booster énormément la méthode de crawl des robots et la qualité d'enregistrement. L'autre mise à jour majeure du moteur a certainement été Google Mayday, déployée un mois avant (mai 2010) et qui a permis de mieux interpréter les mots-clés et les requêtes larges issues de la « longue traîne ».

D'autres mises à jour ont aussi été importantes mais elles relèvent plutôt de l'acquisition ou de la création de nouveaux services, ainsi que de la mise en place de filtres et de pénalités de plus en plus exigeantes. Nous traiterons ce sujet plus en détail par la suite... Nous aurions pu citer également le développement de la recherche en temps réel, notamment autour du *Minty Fresh Indexing* destiné à enregistrer des pages quasiment en instantané afin de toujours proposer les meilleurs résultats aux internautes en temps et en heures.

Continuons notre tour des mises à jour avec le *Knowledge Graph* (ou « graphe de connaissances ») de Google destiné à apporter des informations complémentaires aux usagers lors des recherches. Lancé le 16 mai 2012 aux États-Unis, ce nouvel outil intégré dans les résultats de recherche a pour objectif de comprendre les attentes des internautes et de leur fournir des informations complémentaires sur leur recherche. Le processus se déroule en trois temps :

- l'analyse par Google de la recherche de l'internaute (analyse de chacun des mots-clés de la requête) ;
- la création de relations entre les mots-clés (Google fait appel à une gigantesque base de données pour nouer des liens sémantiques). Fin 2012, cette base de données contenait plus de 500 millions d'entités, ainsi que plus de 3,5 milliards de faits et de relations entre ces différents objets ;

- la proposition de résultats encore plus performants et des informations complémentaires sur la recherche effectuée.

Figure I-3

*La partie à droite est dédiée au Knowledge Graph.*

The screenshot displays a search result for 'Taj Mahal'. On the left, there are several text-based results from various sources like Wikipedia, UNESCO, and travel sites. On the right, the Knowledge Graph is visible, featuring a map of the Taj Mahal, a title 'Taj Mahal', and a list of associated entities such as 'Fort rouge d'Agre', 'Gange', 'Grande Muraille', 'Qutb Minar', and 'Jama Masjid'. Below the graph, there is a section for 'Afficher les résultats pour' with a small portrait of a man, likely related to the 'Musicien' tag.

Le Knowledge Graph de Google analyse le sens des requêtes et tente d'apporter des données supplémentaires pour agrémenter la recherche des internautes. Il est basé sur l'approche ontologique, un modèle de structuration des données pour lequel : un objet de base est une entité, un attribut est une caractéristique de l'objet, une relation est un lien entre plusieurs objets et une classe est un ensemble d'entités.

Il s'agit d'un modèle sémantique et intelligent qui comprend les entités du monde réel et les éventuelles relations qui les lient les unes aux autres. Le Knowledge Graph a donc participé à l'évolution de la recherche sémantique et a apporté une nouvelle vision de la recherche.

Dans les faits, cette évolution marquante de la recherche s'appuie sur des sources publiques telles que Wikipedia, Freebase (abandonnée peu à peu depuis 2015) et CIA World Factbook. Le Knowledge Graph s'applique surtout quand il s'agit de monuments célèbres, de personnalités, de lieux géographiques marquants, de films et musiques, d'œuvres d'art, mais également de sites de marques ou reconnus sur la Toile (mais dans une moindre mesure).



### **Pas d'impact sur le positionnement**

Être dans le Knowledge Graph n'impacte pas directement le positionnement, mais l'effet inverse peut être imaginé. En effet, il est extrêmement rare de voir apparaître le bloc d'informations dans un autre contexte que sur une requête basée sur le nom de domaine d'un site, autrement dit pour une page positionnée en tête des résultats. En définitive, l'intérêt est de renforcer sa visibilité sur des requêtes fortes car les usagers auront du mal à ne pas faire confiance à un site bien positionné mais également présent dans la zone du Knowledge Graph.

Les dernières mises à jour marquantes du moteur de recherche sont certainement Google Hummingbird (2014) et RankBrain (2015), des algorithmes destinés à mieux analyser et comprendre les contenus des pages web et les requêtes des internautes. Nous reviendrons en détail sur ces sujets au début du premier chapitre.

### **Mises à jour de Bing**

Bing est un moteur de recherche assez innovant qui reprend dans les grandes lignes ce que propose Google, ce qui lui a d'ailleurs valu des attaques dans sa courte histoire tant les analogies étaient nombreuses. En pratique, voici les fonctionnalités proposées : recherche d'actualités, de produits (abandonnée en 2013), d'images et de vidéos, Bing Maps (cartographie), recherches associées, historique des recherches, météo en direct, traducteur et calculatrice en ligne, Bing Rewards (un système de crédit présent seulement aux États-Unis) ou encore Bing Voyages.

La technologie de recherche de Microsoft est axée autour d'algorithmes tels que le BrowseRank, créé dès 2008 (source : <http://goo.gl/rdxuqr>), et le StaticRank que nous détaillerons plus tard.


Microsoft communique peu autour de ses mises à jour, sauf si ces dernières sont majeures. Nous allons dresser un rapide historique des mises à jour récentes et marquantes du moteur de recherche.

- Juin 2009 : dans la foulée du lancement de Bing, Microsoft a lancé Bing Travel (ou Bing Voyages) après le rachat du service Farecast en avril 2008 (initialement aux États-Unis).
- Décembre 2009 : Bing Maps est lancé pour contrer Google Maps. Dans les faits, ce type de service existait déjà chez Microsoft depuis 2005 avec Windows Live Local basé sur la technologie Microsoft MapPoint mais le nom de Bing Maps a été attribué en décembre 2009 afin de coller à la nouvelle politique commerciale de la firme. Le service utilise Silverlight de Microsoft et a connu nombre de mises à jour de cartes et de technologies dans son histoire, notamment le 15 décembre 2010 avec une mise à jour graphique de l'outil autour d'un nouveau fond de cartes (après une autre refonte du 7 décembre 2010) et le 12 juin 2013 via l'ajout de 270 To de données.
- Septembre 2010 : Bing Rewards a été annoncé et déployé aux États-Unis afin d'offrir un système de crédits en fonction des recherches des internautes et du temps de diffusion des annonces. Cet ingénieux système économique n'a pourtant pas encore été mis en place partout dans le monde.
- 10 mai 2012 : lancement du Knowledge Graph de Bing appelé Bing Snapshot, soit six jours seulement avant celui de Google (essentiellement déployé aux États-Unis). Le 31 mars 2014, Richard Qian de l'équipe d'indexation et sémantique de Bing a indiqué que plus de 150 millions d'entités ont été ajoutées dans Bing Snapshot (source : <http://goo.gl/pnjuftr>).

Figure I-4

*Bing Snapshot,  
le graph de connaissances  
de Microsoft*

## Winston Churchill



Sir Winston Leonard Spencer-Churchill, KG, OM, CH, TD, DL, FRS, RA was a British politician who was the Prime Minister of the United Kingdom from 1940 to 1945 and again from 1951 to 1955. Widely regarded as one of the greatest wartime lead... +

[en.wikipedia.org](http://en.wikipedia.org)

en.wikipedia.org

Lived: Nov 30, 1874 - Jan 24, 1965 (age 90)

Height: 5' 8" (1.73 m)

Spouse: Clementine Churchill (1908 - 1965)

Children: Randolph Churchill · Mary Soames, Baroness Soames · Diana Churchill · Sarah Churchill · Marigold Churchill

Previous offices: Prime Minister of the United Kingdom (1951 - 1955) · Prime Minister of the United Kingdom (1940 - 1945) +

Parents: Lady Randolph Churchill · Lord Randolph Churchill

---

Listen to audio ^


- Churchill prepares for German invasion of Britain 0:22
- Winston Churchill rallies British citizens 1:35

---


Timeline v

---


Written works




**The River War**  
1899




**A History of the English...**



**My Early Life**  
1930



**While England S...**  
1938



**Savrola**  
1899

- Mi-2013 : Bing déploie son système de produits sponsorisés dans la lignée d'un Google Shopping. Initialement, Microsoft avait déployé son comparateur de prix Bing Shopping (anciennement Microsoft Live Shopping) basé sur Ciao, racheté en 2008, mais ce service a été arrêté en octobre 2013 au profit des produits sponsorisés avec Bing Products Ads, accessibles par l'interface de gestion des liens payants du moteur.
- 11 octobre 2013 : Bing Snapshot accueille désormais les auteurs de contenus grâce au partenariat de Microsoft avec le service Klout (source : <http://goo.gl/2xMlyA>). L'outil permet donc de développer un AuthorShip à la manière de Google, nous détaillerons ce sujet dans les prochains chapitres.
- Mises à jour de l'algorithme de pertinence en mars 2012, fin août 2012, en décembre 2013 et le 5 mai 2014.

- 16 juin 2014 : Microsoft annonce qu'il travaille sur un projet intitulé Bing Catapult (source : <http://goo.gl/sqgC3y>), une nouvelle infrastructure complète des datacenters du moteur de recherche pour accélérer et améliorer l'indexation des pages web et la pertinence des résultats. Le déploiement a été réalisé début 2015 aux États-Unis, puis plus tard dans le monde entier.
- 14 mai 2015 : Bing annonce le déploiement d'un algorithme spécifique à la recherche mobile (source : <http://goo.gl/AI08X3>), un mois après avoir affiché un libellé « mobile-friendly » dans les SERP mobiles.
- 15 juin 2015 : Bing passe totalement en HTTPS avec SSL. Les URL du moteur de recherche sont désormais toutes sécurisées (source : <https://goo.gl/5mSRIX>).

Il est souvent reproché à Bing de copier Google. Bien que cela ne soit pas toujours vrai quand nous comparons les dates de lancement des services, des similitudes sont à déplorer telles que la dernière en date avec *Bing Rich Captions*, un système équivalent aux extraits enrichis de Google que nous détaillerons bientôt...

Figure I-5

*Prix et disponibilité affichés avec Bing Rich Captions*



## Peut-on vivre sans Google ?

Nous considérons souvent que Google est seul au monde, mais il ne faut jamais enterrer les concurrents qui ont parfois un rôle à jouer. Il suffit de suivre les différents baromètres des parts de marché dans le monde pour s'en rendre compte. Google domine, écrase parfois, mais n'est pas toujours prédominant, comme c'est le cas en Russie, en République Tchèque ou encore en Chine.

Aux États-Unis, les parts de visites sont plus réparties que dans d'autres pays bien que Google domine avec 67 % devant près de 19 % pour Bing et 10 % pour Yahoo! (source : <http://goo.gl/HbTEca>).

Si notre marché est francophone, force est de constater que Google domine outrageusement et cela explique en partie pourquoi nous évoquons essentiellement ce moteur de recherche au sein de cet ouvrage. Par ailleurs, Google communique bien plus que Microsoft au sujet de son outil, ce qui ne nous permet pas toujours d'être exhaustifs à propos de Bing.

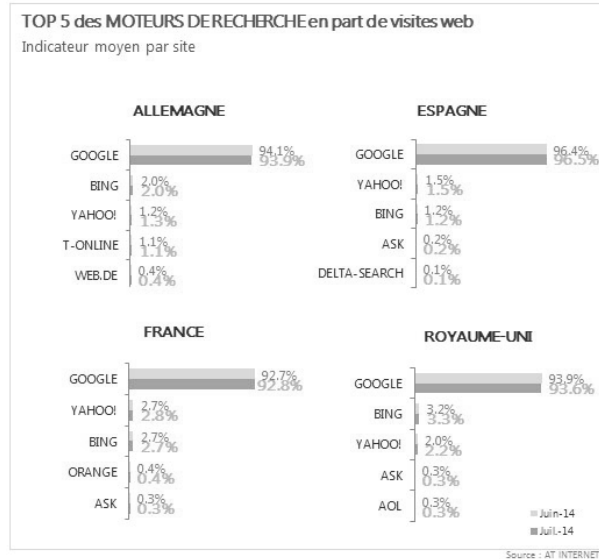
L'institut AT Internet suit l'évolution des parts de marché des moteurs de recherche et montre l'avance considérable de Google sur ses concurrents en Europe occidentale. Nous évoquerons toutefois des moteurs « secondaires » français tels que Qwant ou Exalead pour leurs différences techniques et leurs innovations bien qu'ils ne bénéficient pas d'une place prédominante dans l'esprit des internautes.

Néanmoins, si nous ciblons un marché mondial ou tout du moins marqué par les pays comme la Russie ou la Chine, il est certain que nous devrons mieux maîtriser des moteurs de recherche tels que Yandex ou Baidu notamment. La malice de Microsoft pour conquérir les résultats internationaux de Baidu en

Chine montre que nous devons à tout prix nous focaliser sur ce moteur si nous voulons toucher les visiteurs chinois.

Figure I-6

Répartition des parts de marché des moteurs en Europe



## La programmation au service du référencement

Internet est un monde complexe dans lequel s'affrontent nombre de technologies et de services. Nous pensons souvent maîtriser le Web et ses spécificités, mais sa richesse est telle que nous ne faisons que rarement le tour de la question, et c'est peu dire...

Force est de constater que le référencement fait partie des disciplines en vogue à la portée de tout passionné disposant d'un socle de connaissances suffisant pour administrer des sites web. Malheureusement, trop de spécialistes se voient limités par la barrière du développement et du « code » qui pourrait leur permettre d'aller plus loin en imaginant des programmes adaptés à leurs besoins ou tout simplement en gagnant du temps de gestion. Certes, de nombreux développeurs réalisent cette tâche pour nous, mais cela implique souvent des frais que nous pourrions éviter si nos capacités en programmation étaient plus développées.

L'objectif de cette introduction est de fournir (ou rappeler) des bases en matière de développement, sans pour autant rentrer dans des détails obscurs puisque d'autres livres sont bien plus adaptés pour répondre à ce besoin. Néanmoins, nous allons nous remettre les idées en place voire apprendre des bases afin de pouvoir améliorer notre référencement en conséquence, comme nous le verrons au cours des divers programmes disponibles dans le livre.

## Différencier les langages côté client et côté serveur

La diversité des langages est souvent le premier frein pour les non-initiés. Les choix sont si nombreux en informatique que nous ne savons jamais avec quel langage commencer et surtout pourquoi en utiliser un plutôt qu'un autre. En réalité, la Toile se limite à une quinzaine de langages spécifiques et n'est pas débordée comme la programmation informatique pure qui se compose de plusieurs dizaines d'écritures telles que C++, Ada, Cobol, PacBase, Visual Basic et tant d'autres. Tous ne sont pas obligatoires pour affronter la programmation. Dans cette infime collection, il est important de distinguer les langages initialisés côté client et côté serveur.

Le client correspond au navigateur web, c'est-à-dire l'ordinateur courant, tandis que le serveur est un logiciel installé sur une machine distante (souvent gérée par un hébergeur externe dans le cas des petites structures ou des particuliers).

Globalement, un langage qui s'exécute côté client peut afficher des contenus dans un navigateur et gérer des animations. *A contrario*, un langage côté serveur a une panoplie bien plus développée car il peut notamment s'interfacer avec des bases de données (comme PostgreSQL, MySQL, DB2, SQL Server, Oracle ou encore SQLite) mais aussi gérer des fichiers divers (PDF, CSV...), se connecter à des API (applications web), traiter des formulaires et permettre de créer des interfaces d'administration (*backoffice*).

Les usages sont multiples et cette liste de fonctionnalités est non exhaustive. Il ne faut donc pas se tromper de langages selon les besoins, c'est pourquoi nombre de codes de ce livre seront basés sur des programmes exécutés côté serveur.

Côté client, les langages sont peu nombreux : HTML (et HTML 5) pour la mise en page, CSS pour la mise en forme (CSS 3 actuellement mais le CSS 4 est en développement depuis 2010), JavaScript pour les animations et scripts divers (ou par le biais de bibliothèques telles que jQuery, Rico, Prototype, Mootools...), VBS ou JScript pour être compatible avec les technologies de Microsoft, Flash et enfin les applets Java ou ActiveX (Microsoft).

Côté serveur, la liste peut être plus exhaustive mais si nous nous arrêtons aux langages purement web, ceux-ci sont rares. Dans les faits, plusieurs langages sont issus du monde informatique et ont proposé des modules web avec le temps (comme Python). Nous pouvons citer des systèmes de codage tels que PHP, ASP et ASP.Net de Microsoft, Java (JSP et servlets), ColdFusion, Python, Perl ou Ruby on Rails. Tous reprennent les grandes lignes des langages de programmation mais leur syntaxe et les techniques de développement peuvent totalement varier de l'un à l'autre.

### Microsoft propose des technologies web

ASP et ASP.Net ne sont pas des langages de programmation, ce sont des technologies mises en place par Microsoft. Nous programmons en VBScript (« VBS » ou « Visual Basic Script ») et JScript pour ASP, ou en C# ou VB.Net en ASP.Net. Il faut maîtriser au moins un de ces langages pour coder sur Microsoft.

La société Tiobe<sup>1</sup> dresse la liste des langages de développement informatique préférés des usagers. Sur le Web, le langage qui sort vainqueur côté serveur est PHP tant il est efficace et à la portée de tous. D'autres langages peuvent aussi être utilisés comme VBS ou C# pour les utilisateurs de serveurs Microsoft (IIS). Python est à la mode actuellement, bien que ce ne soit pas un langage web à l'origine (seules des bibliothèques externes le rendent utile sur le Web).

Figure I-7

Liste des langages préférés des développeurs

Dec 2013	Dec 2012	Change	Programming Language	Ratings	Change
1	1		C	17.890%	-0.81%
2	2		Java	17.311%	-0.26%
3	3		Objective-C	10.202%	-0.91%
4	4		C++	8.268%	-0.94%
5	5		C#	5.620%	+0.07%
6	6		PHP	5.281%	-0.26%
7	7		(Visual) Basic	3.752%	-1.42%
8	8		Python	2.210%	-1.64%
9	21	⬆	Transact-SQL	1.877%	+1.30%
10	11	⬆	JavaScript	1.852%	+0.53%
11	15	⬆	Visual Basic .NET	1.688%	+0.80%
12	9	⬇	Perl	1.072%	-1.10%
13	10	⬇	Ruby	0.932%	-0.80%
14	17	⬆	MATLAB	0.708%	+0.10%
15	12	⬇	Delphi/Object Pascal	0.691%	-0.29%

## Quel rôle pour le référencement ?

Connaître les différents langages, leur utilité et leur degré d'intérêt permet de sélectionner les technologies adaptées à chaque besoin. En référencement, il est important de distinguer tous ces langages car certains d'entre eux sont impitoyables et peuvent engendrer des pertes de trafic importantes. Il convient de rester méfiant pour toujours trouver le bon langage et le code idéal pour chaque usage.

Tout au long de notre périple, nous allons être confrontés à une multitude d'outils avec des méthodes de conception parfois bien différentes. En effet, nous n'utilisons pas un moteur de blog tels que WordPress ou DotClear comme un CMS tels que Drupal, Spip ou Django (en Python) ou encore un framework comme eZpublish, Symfony, CakePHP, Zend ou Play! (en Java et Scala). Il faut bien maîtriser le code pour se lancer dans l'utilisation de certains outils, c'est l'une des raisons qui explique le succès de WordPress, par exemple, tant son utilisation avancée reste accessible.

1. <http://www.tiobe.com/index.php/content/paperinfo/tpci/index.html>

Lorsque les référenceurs doivent se plonger dans les codes pour optimiser les contenus, ils se limitent souvent aux balises HTML car cela reste le fondement essentiel de ce métier. Toutefois, il est parfois nécessaire d'aller plus loin pour automatiser des tâches, ou tout simplement pour gérer des contenus profonds dissimulés dans des morceaux de code écrits avec d'autres langages. Dans ce cas, il est important d'avoir au moins les bases de développement pour se repérer mais aussi pour créer nos propres fonctions utiles pour le référencement. C'est tout l'objet de notre lecture...

Afin que tout soit bien clair après cette introduction, sachez que nous ne rappellerons que le B.A.-ba de la programmation dans peu de langages et que nous ne proposerons pas de codes orientés objet car cette syntaxe, aussi qualitative soit-elle, n'est pas encore la plus courante (les développeurs les plus aguerris et jeunes se tournent vers cette syntaxe mais ce phénomène n'est pas généralisé).

Par conséquent, nous ne traiterons que des bases HTML-CSS, en PHP et JavaScript mais nous négligerons le reste, bien que ce livre présente des codes fondés sur la technologie ASP de Microsoft. Nous devons faire des choix et réduire ces explications à quelques lignes auraient peu de valeur, nous préférons que vous vous orientiez vers des ouvrages exhaustifs sur ces sujets si vous craignez de ne pas comprendre tous les programmes commentés présents tout au long de notre propos.

#### **Différence entre développement procédural et POO**

Il faut savoir qu'il n'existe pas de réelles différences en matière de performance selon que nous codons avec la méthode procédurale ou orientée objet (POO), contrairement à ce que pléthore de développeurs avancent pour vanter les mérites de telle ou telle technique. Chacun est libre de coder comme il le souhaite tant que le résultat est fonctionnel, compréhensible (bien commenté notamment) et réutilisable par un tiers.

WordPress est un bon exemple car son code Open Source est contrebalancé entre la programmation procédurale et orientée objet selon les développeurs qui interviennent à la source. Nous trouvons même des morceaux de codes en PHP 4 et en PHP 5 mélangés. Les bases sont essentielles pour ne pas être perdu...

Connaître la programmation web a toujours fait débat dans le milieu du référencement. D'un côté, les réfractaires se refusent à coder la moindre ligne car ils estiment que ce n'est pas toujours utile et que des spécialistes peuvent le faire à leur place. D'un autre côté, au contraire, les partisans voient dans la maîtrise du code un plus non négligeable et surtout un moyen de mieux gérer leur référencement.

Force est de constater qu'un non-développeur ne peut pas tout faire en matière de référencement tant ses limites techniques l'empêchent d'imaginer des programmes automatisés ou même de répondre à certains besoins complexes. En effet, comment gérer le fichier `.htaccess` pour optimiser le référencement si nous ne savons rien coder ? Comment améliorer un système multilingue si nous ne différencions pas la bonne méthode des mauvaises techniques ? Ce ne sont que des exemples, mais la liste pourrait facilement s'étendre si nous grattions un peu le sujet...

Désormais, les référenceurs sont nombreux sur le marché du travail et nombre d'agences préfèrent trouver la perle rare voire le couteau suisse plutôt que d'avoir un référenceur spécialiste uniquement de l'aspect webmarketing. Dans ce cas, avoir de bonnes connaissances techniques permet de sortir du lot et de trouver un poste plus polyvalent avec des possibilités accrues en matière d'optimisation.

L'autre problématique vient de la taille réduite des entreprises et des start-ups relatives à Internet. Nous trouvons beaucoup de PME voire de TPE dans ce secteur, et nous n'avons donc pas toujours des développeurs à portée de main pour coder à notre place. Il faut par conséquent mettre les mains dans le cambouis pour trouver des solutions.

Au fond, souhaitons-nous être éternellement limité et dépendant d'autres personnes ou préférons-nous contrôler le référencement de A à Z ? Bien entendu, il n'est pas utile d'être un génie du développement ou de tout maîtriser. D'une part, cela est impossible et d'autre part, seulement une infime partie serait directement utile pour gagner du temps et optimiser nos tâches. Malgré tout, retenons qu'une bonne maîtrise technique confère une réelle indépendance et une liberté dans l'exécution du métier, c'est pourquoi nous verrons diverses applications SEO dans la suite de notre lecture.

## Socle technique pour programmer sur le Web

### Langage HTML et mise en page

HTML est un langage de balisage commun qui se décompose en deux parties encadrées par des balises `<html>` et `</html>`.

- La « tête », balisée par `<head>` et `</head>`, contient les informations nécessaires au bon fonctionnement de la page, dont le titre de la page si cher au référencement ou encore la balise permettant de relier le document à des styles CSS.
- Le « corps » est encadré par les balises `<body>` et `</body>` et contient toutes les données visibles par les utilisateurs, à savoir les contenus, les images, etc.

Ces deux parties insérées entre les balises `<html>` et `</html>` sont surmontées d'un doctype, c'est-à-dire une déclaration de type de document qui permet d'indiquer au navigateur et aux robots quelle version du langage est utilisée. Par exemple, si vous choisissez HTML 5, le doctype est simpliste :

```
<!DOCTYPE html>
```

Une fois cette structure initiale réalisée, vous pouvez démarrer en remplissant la tête du document avec les éléments essentiels (jeu de caractères à utiliser, titre et description du document, mots-clés associés, lien vers les styles CSS, scripts JavaScript...).

#### Quel charset utiliser ?

Le choix du jeu de caractères est primordial en développement web. Les Européens sont toujours partagés entre l'ISO-8859-1 (caractères occidentaux) et l'UTF-8 (tous types de caractères), mais il est fortement conseillé d'opter pour le second choix, bien plus polyvalent et compatible à l'international.

Voici à quoi peut ressembler notre début de page web en HTML :

```
<!DOCTYPE HTML>
<html>
<head>
```



```
<meta charset="utf-8"/>
<title>Titre du document</title>
<meta name="description" content="Description du document"/>
<meta name="keywords" content="mot-clé 1, mot-clé 2, ..."/>
<link rel="stylesheet" type="text/css" href="style.css"/>
<script type="text/JavaScript">// Scripts potentiels...</script>
</head>
<body>
Contenu de la page (visible par les utilisateurs)
</body>
</html>
```

Il faut ensuite se concentrer sur le corps de la page qui contient les contenus des pages web à l'aide de balises « block » (qui forcent un retour à la ligne et prennent des dimensions si nécessaire) et « inline » (sans retour à la ligne ni largeur et hauteur configurables) plus ou moins nombreuses selon la version HTML utilisée.

Les balises de type « block » les plus connues sont `<div>` (bloc de contenu), `<h1>` (jusqu'à `<h6>`) et `<p>` (paragraphe de texte) tandis que les balises en ligne sont plus nombreuses, telles que `<span>` (balise pour styliser des parties de contenus), `<strong>` (mise en exergue représentée par une mise en gras), `<em>` (mise en avant avec affichage en italique) ou encore `<a>` (liens hypertextes).

#### Quel rôle pour le référencement naturel ?

La quasi-totalité de ces balises a un impact sur le référencement naturel, les mots-clés contenus entre ces éléments ont plus de valeur, à l'exception des éléments `<p>` et `<div>` qui sont « neutres ».

En HTML 5, nous trouvons davantage de balises structurales telles que `<header>` (en-tête), `<footer>` (pied de page), `<nav>` (zone de navigation), `<aside>` (contenus annexes), `<hgroup>` (groupements de titres), `<section>` et `<article>` pour les contenus. Cette mouture offre plus de précision sémantique et permet de mieux hiérarchiser les textes. Nul doute que cela influencera le référencement et le positionnement des pages dans un avenir proche.

Enfin, avant de donner un exemple complet d'un corps de page, il faut savoir que les menus sont souvent réalisés à l'aide de balises de listes (`<ul>...</ul>`) dans lesquels chaque item de liste (`<li>...</li>`) contient un lien hypertexte.

```
<body>
<div id="bloc-general">
  <header>
    <h1>Logo du site</h1>
  </header>
  <nav>
    <ul>
      <li><a href="page1.html">Page 1</a></li>
      <li><a href="page2.html">Page 2</a></li>
      <li><a href="page3.html">Page 3</a></li>
```

```
</ul>
</nav>
<section>
  <article>
    Texte avec <strong>mise en gras</strong> et <em>mise en italique</em>.
  </article>
</section>
<footer>
  <p>Pied de page</p>
</footer>
</div>
</body>
```

Nous n'avons pas pu traiter l'ensemble des spécificités des versions d'HTML ici. Vous pourrez également rencontrer des balises de formulaire (comme `<form>`, `<input />...`) ou de tableaux (comme `<table>` ou encore `<tr>` et `<td>`), par exemple. Il est fortement conseillé de maîtriser ce langage tant il est primordial pour le référencement. Son compère, le CSS, est utile pour réaliser la mise en forme de vos pages web mais il ne joue pas de rôle majeur en SEO.

## Gérer les animations avec JavaScript, jQuery, Prototype et consorts...

Dans la lignée des langages exécutés côté client, JavaScript et toute sa panoplie de bibliothèques sont à connaître pour donner vie aux pages web tant les animations et les interactions générées par ces codes peuvent apporter un plus pour les utilisateurs. Néanmoins, il est important de garder en mémoire que le JavaScript et ses acolytes ne sont pas les plus grands amis de Google car une grande partie des codes sont bloquants pour les robots du moteur de recherche. Il convient donc de l'utiliser avec parcimonie.

Généralement, la programmation est fondée sur un principe simple : l'usage de « variables », à savoir des entités créées pour être réutilisées comme bon nous semble en cas de besoin. Les variables peuvent recevoir une valeur fixe ou dynamique, ce qui leur permet d'interagir avec les programmes en fonction des données captées ou directement proposées par l'utilisateur.

En JavaScript, tout comme dans les bibliothèques associées, une variable est définie en ajoutant le mot-clé `var` avant son nom (`var uneVariable`). Une variable peut prendre plusieurs types de valeurs, par exemple une chaîne de caractères, un numérique ou encore un booléen (`true/false`).

Le deuxième point qui a fait le succès de la programmation est le système conditionnel qui permet d'effectuer des actions précises en fonction d'hypothèses fixées au préalable. Ce mécanisme fonctionne avec la syntaxe `if (condition) {...} else {...}`.

L'autre facteur essentiel à comprendre lorsque nous développons sur le Web est le système des boucles, dont l'objectif est de pouvoir répéter des actions autant de fois que nous le souhaitons. Trois types de boucles se retrouvent dans la majorité des langages :

- `while` signifie « pendant que la condition est vraie, fais ceci » et s'écrit `while (condition) {...}` ;

- `do ... while` est à peu près équivalente, à la seule différence que le premier tour de boucle est réalisé, même si la condition n'est pas respectée à la fin du tour. Sa syntaxe est `do {...} while (condition);`
- `for` permet d'ajouter une incrémentation automatique et de parcourir des tableaux de données notamment. Elle s'écrit sous la forme `for (variable = valeur; condition; incrémentation) {...}`.

Hormis ces grandes règles de programmation, la force de JavaScript et des bibliothèques telles que jQuery ou Prototype (qui facilitent la vie des développeurs) est de pouvoir interagir avec les éléments HTML ou avec les propriétés CSS pour créer des animations ou modifier des éléments. Nous croiserons d'ailleurs quelques scripts au cours de notre lecture qui iront dans ce sens.

La fonction `getElementById()` ; résume la méthode la plus courante pour capter un contenu selon son identifiant unique. Voici un exemple de code en HTML et JavaScript :

```
<div id="bloc1">Texte écrit en vert avec CSS</div>
<a href="#" onclick="document.getElementById('bloc1').style.color = blue ;">Changer la
couleur du texte en bleu</a>
```

jQuery et les autres bibliothèques JavaScript élargissent encore davantage les possibilités tout en simplifiant l'écriture du code pour la majorité des modifications courantes. Par exemple, le changement de couleur s'écrirait ainsi en jQuery :

```
$("#bloc1").css('color','blue');
```

Enfin, le dernier concept à connaître avant de continuer notre parcours rapide des langages web est celui des fonctions. En programmation, tout ou presque n'est que fonction, c'est-à-dire des mini programmes qui permettent de gagner du temps et d'éviter de programmer plusieurs fois les mêmes scripts. La syntaxe est toujours de la forme suivante :

```
function NomDeLaFonction(arguments-optionnels) {
// code de la fonction
}
```

Il ne s'agit que de bases à approfondir mais elles vous donnent la teneur de la programmation en JavaScript natif ou à l'aide des bibliothèques (notamment jQuery). Terminons notre rapide introduction par PHP.

## Bases de PHP

Le langage PHP est l'un des plus répandus sur la Toile tant il est puissant et permet d'effectuer une multitude de tâches côté serveur. Nous n'allons pas beaucoup développer ce langage puisque les mêmes principes que ceux cités précédemment s'y retrouvent, à savoir les variables, les boucles, les conditions mais aussi la création des fonctions.

### Baliser les codes PHP

Les codes PHP sont toujours encadrés par des balises `<?php ... ?>` et se repèrent très vite dans le code.

En réalité, la différence majeure de syntaxe entre JavaScript et PHP réside dans les variables qui ne se basent pas sur le mot-clé `var` mais tout simplement avec le signe du dollar `$`. Ainsi, il suffit d'écrire `$uneVariable` pour créer une variable de notre choix. Pour le reste, les boucles conservent la même syntaxe, tout comme les systèmes conditionnels et les fonctions.

#### Avantage des fonctions en PHP

Les fonctions PHP présentent un avantage car elles peuvent accueillir des valeurs par défaut dans les paramètres de fonctions, ce qui s'avère très pratique.

Un autre aspect majeur de PHP concerne le traitement des formulaires puisque nous utilisons sans cesse des champs ou des zones de texte pour interagir avec les utilisateurs (commentaires, formulaires de contact, backoffice...), c'est pourquoi une bonne maîtrise des types de données et de leur traitement est primordiale.

Figure I-8

Exemple de méthode `GET` avec le paramètre « s »



Pour récupérer les informations issues des champs de formulaire, PHP met à disposition deux variables dites « super-globales » : `$_GET['name']` et `$_POST['name']`. Entre les crochets, il suffit d'indiquer la valeur de l'attribut `name` des champs de formulaire en HTML pour récupérer la valeur (`value`). Le choix entre `GET` et `POST` correspond à l'attribut `method` précisé dans la balise `<form>` en HTML.

#### Différences entre `GET` et `POST`

`GET` fait passer des paramètres dans les URI tandis que `POST` dissimule les informations. Nous préférons `GET` pour tout ce qui contient une pagination en général (galeries, moteur de recherche...) et la méthode `POST` lorsque nous envoyons des messages ou que nous passons des données sensibles (module d'identification avec mot de passe, formulaire de contact...).

```
<form method="get" action="">
  <input type="text" name="recherche" />
  <input type="submit" name="bouton" value="Rechercher" />
</form>
<?php
// Si le bouton est cliqué
if(isset($_GET['bouton'])) {
  // On affiche la requête saisie
  echo "Requête de recherche : ".$_GET['recherche'];
}
?>
```

Les formulaires présentent de nombreux risques en matière de sécurité. Il faut veiller à bien protéger les variables et leurs données dans la grande majorité des cas avec des fonctions spécifiques telles que `htmlspecialchars()`, `isnumeric()`, `addslashes()`, etc.

Enfin, terminons notre tour d'horizon des concepts avec celui des tableaux, disponibles aussi bien en PHP qu'en JavaScript, par exemple. Les tableaux jouent un rôle majeur car nous les utilisons sans cesse, notamment lorsque nous parcourons des bases de données.

En PHP, nous créons les tableaux « scalaires » en faisant appel à la fonction `array()`. Les items des tableaux sont séparés par des virgules comme dans l'exemple suivant :

```
$tableau = array('item1', 'item2', 'item3', '...');
```

#### Et les tableaux en JavaScript ?

En JavaScript, le principe est identique mais nous devons ajouter le mot-clé `new` pour créer l'objet `tableau` comme dans `var tableau = new Array('item1', 'item2...');`

Lorsqu'il s'agit de tableaux simples comme celui présenté ici, nous faisons appel aux valeurs en les ciblant à l'aide de leur « identifiant » dans le tableau. Le premier item est toujours « 0 », le second « 1 », et ainsi de suite. Si nous voulons récupérer la valeur "item2", par exemple, le fait qu'elle soit en seconde position dans le tableau implique que nous devons écrire ce qui suit :

```
echo $tableau[1];
```

Il existe également des tableaux dits « associatifs » qui permettent de personnaliser les clés reliées à des valeurs, plutôt que d'être limité à des numéros sans lien logique avec les données. Dans ce cas, les items sont toujours séparés par des virgules, mais les clés et les valeurs sont distinguées par une « flèche », comme ici :

```
$tableau = array('cle1'=>'valeur1', 'cle2'=>'valeur2', '...'=>'...');
```

Les bases de données ou tableaux complexes se présentent comme des tableaux placés dans des tableaux. Ils sont appelés « multidimensionnels » et peuvent avoir des tailles démesurées.

```
$tableaumulti = array(
    0 => array('nom'=>'Martin', 'prenom'=>'Alexandra'),
    1 => array('nom'=>'Chartier', 'prenom'=>'Mathieu'),
    2 => array('nom'=>'Andrieu', 'prenom'=>'Olivier')
);
echo $tableaumulti[0]['prenom']; // Affiche Alexandra
```

Enfin, sachez qu'il existe une boucle particulière (`foreach`) qui permet de parcourir les tableaux rapidement. Elle se traduit par : « pour chaque item du tableau, parcourir les données ». Sa syntaxe est un peu particulière, comme le montre les exemples suivants :

```
$tab = array('nom'=>'Chartier', 'prenom'=>'Mathieu');  
// Syntaxe : foreach($nom-tableau as $valeur-tableau) { ... }  
foreach($tab as $valeur) {  
    echo $valeur; // Affiche "ChartierMathieu"  
}  
// 2e écriture : foreach($nom-tab as $cle-tab => $valeur-tab) { ... }  
foreach($tab as $cle => $valeur) {  
    echo $cle; // Affiche "nom" puis "prenom"  
    echo " : "; // Sépare les clés des valeurs par " : "  
    echo $valeur; // Affiche "Chartier" puis "Mathieu"  
}
```

Nous n'avons pas pu traiter tous les concepts intéressants de PHP ou de JavaScript tels que la concaténation, l'arithmétique ou encore la gestion des objets mais nous vous invitons à vraiment intégrer ces concepts pour progresser. Il est fortement conseillé de se référer à la documentation officielle de PHP pour aller plus loin avec ce langage (<http://www.php.net/manual/fr/>).

## Conclusion sur la programmation

D'autres langages peuvent être croisés lorsque nous faisons du Web tels que le Python avec son écriture très segmentée et imbriquée, ses `if... elif...` ou encore ses tuples et listes de données. Java est un langage orienté objet courant et très typé également, il peut se croiser de temps en temps, tout comme les technologies ASP voire ASP.Net si vous travaillez avec des environnements Microsoft.

Nous ne pouvons pas traiter de tous ces langages en quelques lignes, ce n'est pas le sujet du livre. Comme nous l'avons dit au début de ce chapitre, l'objectif est d'utiliser les forces de la programmation au profit du référencement, et bien que cela demande une mise à niveau ou quelques rappels, chacun sera libre d'approfondir les langages cités précédemment.

Retenons que nous devons maîtriser HTML pour gérer notre référencement et que tous les autres langages qui gravitent autour de lui pourront avoir un impact indirect sur notre travail, soit en améliorant l'ergonomie générale des sites web avec CSS, JavaScript ou même PHP, soit en nous permettant de créer nos propres outils pour automatiser des tâches et gagner en cadence. Une fois les techniques acquises, nos seules réelles limites sont notre imagination et notre créativité, sans quoi nous pourrions développer une multitude d'outils utiles pour les référenceurs.

Enfin, n'oublions pas non plus la suite Office de Microsoft et VBA (ou VBE pour Excel) qui peuvent aussi nous permettre de créer des macros puissantes pour classer et administrer des données ou encore générer des rapports, bien que cela soit indépendant du Web mais aide surtout le travail en amont et en aval du référenceur.

Nous allons désormais étudier plusieurs sujets connus dans le milieu SEO en allant dans les détails et en tentant de créer des programmes clés en main (dont certains peuvent être améliorés) pour nous aider dans nos tâches rébarbatives ou complexes.