

Comment réaliser un audit du SEO technique

HubSpot OnCrawl



Table des matières

1	Introduction
4	Les objectifs d'un audit technique
10	Les outils essentiels
14	L'indexation
21	Les performances des serveurs
29	La structure d'un site
39	Les performances des mots-clés
45	L'optimisation on-page
52	Le contenu dupliqué
56	Les problématiques liées aux appareils mobiles
60	Les problématiques liées aux sites internationaux
65	Schema.org
71	Tirer les conclusions d'un audit
73	Conclusion

Introduction



Introduction

Comprendre les mécanismes du SEO technique d'un site est essentiel au succès de ses activités en ligne. Toutefois, rester au fait des nombreuses modifications que Google apporte à son algorithme relève parfois du défi, même pour des marketeurs à la pointe de la technologie. HubSpot et OnCrawl se sont donc associés pour créer ce guide sur l'audit du SEO technique, afin d'aider les entreprises à exploiter le plein potentiel de leur site web. Cet e-book rassemble les outils, les informations et les stratégies indispensables pour faire du SEO technique l'un des piliers de l'optimisation des performances d'un site web.

Le SEO, ou l'optimisation pour les moteurs de recherche, fait référence aux techniques qui permettent à un site d'obtenir un meilleur classement sur les pages de résultats des moteurs de recherche. Le SEO technique, qui constitue l'une de ses ramifications, inclut des facteurs clés pour le référencement qui sont intégrés à la structure du site, comme l'exploration et l'indexation. Pour définir l'impact du SEO technique sur les performances globales d'un site, il est donc nécessaire d'effectuer un audit technique.

Qu'est-ce qu'un audit technique ?

Il s'agit du processus d'analyse et de suivi d'indicateurs de performance clés qui s'applique à l'ensemble d'un site web. Il vise à identifier les faiblesses structurelles qui affectent ses performances, son SEO et son classement dans les résultats de recherche.

Qui doit effectuer cet audit ?

La nature technique de ce type d'audit pouvant être dissuasive, ce guide fournit aux marketeurs de contenu, aux développeurs web, aux analystes et à tous ceux qui souhaitent réaliser un audit SEO l'ensemble des outils et des informations indispensables à son succès. Un audit technique permettant d'identifier les leviers d'amélioration d'un site, il implique généralement une collaboration avec l'équipe web pour définir des solutions concrètes.

Quand un audit technique doit-il être réalisé ?

Un audit technique est souvent considéré à tort comme une tâche ponctuelle, alors qu'il doit être envisagé comme un projet à long terme permettant d'évaluer régulièrement les performances d'un site web. Parce que celui-ci évolue et s'étoffe à mesure que l'entreprise se développe, il faut pouvoir être capable de définir les répercussions de ces évolutions sur les performances de la stratégie SEO.

Voici quelques exemples de moments propices à la réalisation d'un audit du SEO technique :

- Peu de temps après la création d'un site
- Suite à une mise à jour de son design
- Suite à une mise à jour de sa structure
- Lorsque le dernier audit remonte à plus d'un an

Pourquoi HubSpot et OnCrawl se sont-ils penchés sur le sujet de l'audit technique ?

HubSpot propose une suite complète de logiciels pour le marketing, la vente et le service client, qui s'articulent tous autour d'un CRM gratuit. Faisant partie des leaders d'opinion dans l'univers numérique, la société s'est spécialisée dans la croissance des entreprises en développant la méthodologie inbound.

OnCrawl est une plateforme dédiée au SEO technique, leader de la mise en place d'audits SEO et de monitoring quotidien pour les entreprises. Elle propose des solutions qui permettent à des entreprises du monde entier de mieux comprendre les principes de référencement de Google et la technicité de leur propre site web, afin d'en améliorer les performances.

Ensemble, HubSpot et OnCrawl offrent leur expertise aux entreprises qui souhaitent optimiser leurs performances.

Les objectifs d'un audit technique



Les objectifs d'un audit technique

Rechercher les erreurs techniques

Les erreurs techniques révèlent certains des problèmes qui desservent les performances d'un site, comme la vitesse de chargement des pages, le nombre de liens cassés ou la redondance du contenu. Il est donc crucial de les résoudre pour optimiser les performances d'une stratégie SEO et l'expérience des utilisateurs.

Le SEO a pour objectif d'optimiser le trafic naturel vers un site, c'est-à-dire le nombre de visites à partir des pages de résultats des moteurs de recherche. Aujourd'hui, face à une concurrence toujours grandissante, cela implique de proposer un site qui soit source de valeur ajoutée pour ses visiteurs. Les problèmes liés au temps de chargement des pages, par exemple, peuvent entraîner une diminution du trafic et nuire à l'expérience utilisateur.

Améliorer le classement d'un site sur les pages de résultats des moteurs de recherche

Comprendre et s'adapter aux réponses fournies par Google

RankBrain est le système d'intelligence artificielle de Google qui traite et évalue les résultats de recherche. Il permet à la plateforme de mieux interpréter les requêtes de ses utilisateurs et de mieux identifier les sites web source de valeur ajoutée pour ces derniers. Pour cela, RankBrain s'appuie sur l'apprentissage automatique pour proposer des pages de plus en plus pertinentes dans les résultats de recherche, à mesure qu'il analyse un nombre croissant de requêtes quotidiennes.

Les performances de RankBrain s'améliorant avec chaque recherche, développer une stratégie SEO pérenne présente un certain nombre de défis. Pour y faire face, une solution consiste à se focaliser sur le SEO technique et sur les mots-clés de longue et moyenne traîne, c'est-à-dire des mots-clés qui ne sont ni trop génériques, ni trop peu ciblés. Le SEO technique permet à Google d'identifier et de comprendre un site plus facilement, tandis que les mots-clés sont essentiels à la stratégie de création de contenu. Des mots-clés définis avec soin favorisent en effet le classement d'une page dans les résultats de recherche pour mieux se démarquer de la concurrence.

Aider Google à définir l'importance des pages clés d'un site

Page Importance est l'algorithme de Google qui classe les URL en les évaluant lors de l'exploration. Il existe à ce jour près de 2 milliards de pages web, que les robots d'exploration ne sont tout simplement pas en mesure de parcourir en totalité. Google s'appuie sur Page Importance pour mieux comprendre quelles pages ces robots doivent cibler en priorité.

De nombreux facteurs entrent en ligne de compte lorsque l'algorithme de Google détermine **l'importance d'une page** :

- La position de la page
- Le classement de la page
- Le type de page ou de fichier
- Le fichier sitemap.xml
- Le classement interne de la page
- Le nombre et le type de liens internes
- Le nombre de liens externes
- La pertinence et la qualité du contenu
- La fréquence des mises à jour
- La qualité du code source et du site web

En optimisant tous ces facteurs, une entreprise peut accroître l'importance de ses pages web et aider Google à identifier celles qui doivent être explorées et analysées en priorité sur son site.

Optimiser les dépenses prévues au budget d'exploration

Le budget d'exploration correspond au nombre de pages d'un site web explorées ou visitées par les moteurs de recherche au cours d'une période donnée. Les moteurs de recherche se contentant d'indexer les pages qu'ils découvrent lorsqu'ils parcourent un site, c'est-à-dire de les rendre disponibles comme résultats de recherche, il est crucial d'optimiser ce budget.

Il est possible d'améliorer le référencement naturel d'un site en travaillant son contenu de sorte que Google explore un plus grand nombre de ses pages clés tout en consacrant un budget plus restreint aux pages moins importantes.

Pour une entreprise, le budget d'exploration correspond au nombre moyen de visites quotidiennes d'un robot d'exploration de Google sur son site. Il est possible de le calculer en analysant les journaux du serveur du site ou d'en obtenir une estimation dans les statistiques d'exploration de la Search Console de Google.

Comment évaluer le budget d'exploration :

1. Établir le nombre de pages du site, qu'elles soient ou non indexées par Google, à l'aide d'un outil comme OnCrawl, qui permet de combiner des données d'exploration et des données issues des fichiers sitemap ou de la Search Console de Google.

Pour obtenir une estimation de ce nombre, il suffit d'effectuer la recherche suivante dans Google : « site:votresiteweb.com ». Le nombre total de résultats indique le nombre de pages indexées du site, qui sera inférieur au nombre total de pages qu'il possède.

2. Accéder aux statistiques d'exploration de la Search Console de Google.
3. Diviser le nombre total d'URL du site par le nombre moyen de pages explorées par jour et définir un budget qui permette d'explorer l'intégralité du site en moins de 5 jours.

Les entreprises n'ont pas la main mise sur le budget d'exploration alloué à leur site. Google en a la charge et l'augmente, selon ses propres termes, en fonction de la capacité de leurs pages à répondre aux requêtes. Par conséquent, tirer profit de ce budget passe non seulement par la mise au point de la stratégie SEO globale des entreprises, mais aussi par l'optimisation du budget dont elles disposent.

S'approprier les positions clés dans les résultats de recherche pour augmenter le trafic

Tous les résultats de recherche ne se valent pas. Certains génèrent davantage de clics, d'autres s'accompagnent d'informations supplémentaires, d'autres encore correspondent au seul résultat disponible pour une recherche vocale, ou au seul contenu consulté par des utilisateurs à la recherche d'une réponse rapide et informative.

« Chercher une réponse rapide et informative » est un exemple d'intention de recherche, soit l'objectif ultime de la personne qui formule une requête ou l'information qu'elle cherche à obtenir. Une intention de recherche peut être de nature informative (Combien font 14 mph en km/h ?), navigationnelle (Comment accéder à la page des réclamations sur le site de telle compagnie aérienne ?), ou transactionnelle (Acheter un manteau d'hiver en ligne).

En tenant compte de l'intention de recherche lorsqu'elle définit les mots-clés ciblés dans sa stratégie, une entreprise peut généralement améliorer le classement des pages de son site, voire le faire apparaître avant les résultats de recherche.

Cette position, que l'on appelle Position 0 ou **extrait optimisé**, est réservée aux résultats si pertinents pour une recherche que Google les affiche en premier, au-dessus de tous les autres. Améliorer les performances d'un contenu pour qu'il soit sélectionné comme extrait optimisé implique qu'il fournisse une réponse à une question précise sur une page dont le SEO technique a été optimisé. En voici un exemple que l'on pourrait trouver sur un blog, une page de destination ou d'autres pages d'un site web :



Vise à être l'extrait optimisé pour la recherche « inbound marketing/qu'est-ce que l'inbound marketing » en incluant le terme exact recherché et en le définissant dès les premières lignes de la page.

Lorsque Google identifie un contenu comme la réponse la plus pertinente pour une recherche, son algorithme peut le sélectionner comme extrait optimisé, autrement dit lui attribuer la Position 0.

Cette dernière peut également être occupée par un carrousel d'images composé d'une série de fiches présentant différents éléments, comme des recettes, des actualités, des photos et plus encore. Ces éléments sont issus des données structurées d'un site web.

Les données structurées aident Google à créer des extraits descriptifs plus informatifs dans ses résultats de recherche, ce qui incite les utilisateurs à cliquer sur une page. Ces résultats peuvent comprendre un classement par étoiles, des liens vers le site et même des vignettes d'images.

Oncrawl SEO page audit - Chrome Web Store

<https://chrome.google.com/webstore/.../oncrawl.../ngcfjkiphejojdjnmiegbcjhemlbdjhp...> ▼

★★★★★ Rating: 5 - 4 votes - Free - Chrome

Jan 22, 2018 - The **Oncrawl** extension provides valuable informations on the status of any web page by measuring its content (SEO tags, duplicate and near ...

Extrait informatif contenant un classement, le tarif (gratuit) et le navigateur compatible recommandé.

Cibler les problèmes techniques qui nuisent à l'expérience utilisateur

L'expérience utilisateur est l'un des éléments de la conception de site web devant entrer en ligne de compte dans la réalisation d'un audit technique.

L'emplacement des pages les plus rentables

Les pages les plus performantes, celles qui ont pour but d'attirer le plus de visiteurs, doivent être placées de façon stratégique dans la structure d'un site. La page d'achat d'un produit ou un article de blog de qualité, par exemple, ne doivent pas se retrouver enfouis dans les profondeurs de son architecture. Il est donc essentiel de définir ce que constitue un contenu de qualité et d'en faciliter l'accès aux utilisateurs, tout en répondant aux critères des robots d'exploration de Google.

L'accessibilité

L'expérience client peut également interférer avec d'autres éléments clés participant au classement d'un site, comme la vitesse de chargement des pages, leur taille, l'optimisation pour les appareils mobiles ou la localisation du contenu. Les pages difficiles d'accès pour les utilisateurs desservent aussi le travail des robots d'exploration lorsqu'ils cherchent à les consulter, à les lire ou à les interpréter. Les problèmes d'accessibilité d'un site se situent au niveau de sa conception et entraînent des conditions de navigation difficiles pour certains groupes d'utilisateurs, comme les utilisateurs d'appareils mobiles, ceux dont la connexion est lente, ceux qui ont recours aux aides visuelles ou encore ceux qui résident à l'étranger.

Il est donc essentiel que la conception d'un site soit de qualité, tant au niveau visuel que structurel, pour qu'il bénéficie d'un meilleur classement dans les résultats de recherche. Le temps de chargement des pages doit être bref, la taille des fichiers aussi réduite que possible et le design, adaptatif. Le site devrait également être disponible en plusieurs langues afin de cibler des audiences locales et internationales. Cette option de localisation, généralement disponible en haut ou en bas de page, permet aux utilisateurs et aux robots d'exploration de mieux comprendre la structure du site et d'accéder aux sections qui les intéressent plus facilement.

Être attentif aux objectifs commerciaux

Les entreprises doivent prioriser les améliorations techniques les plus rentables pour leur activité, qu'il s'agisse d'augmenter le volume de trafic vers leur site web, de générer davantage de leads pour leur équipe commerciale ou de réaliser plus de transactions en ligne. Par exemple, un restaurant local n'aura pas besoin d'investir dans un site international, mais il devra s'assurer que le temps de chargement de ses pages sur appareil mobile et ses données structurées sont optimisés, afin que les clients potentiels de passage puissent accéder facilement au site, au moment opportun.

Quels que soient ses objectifs commerciaux, une entreprise doit les aligner sur sa stratégie SEO pour les atteindre.

Les outils essentiels



Les outils essentiels

Il est impossible de réaliser un audit technique sans les outils appropriés. La liste ci-dessous en référence quelques-uns à utiliser de pair avec cet e-book.

1. **OnCrawl**. Pour comprendre comment les robots d'exploration de Google indexent son site, une entreprise doit d'abord utiliser son propre robot. Associé aux données des fichiers journaux de son serveur, ce robot permet d'obtenir une vue d'ensemble complète et précise de l'activité du site. OnCrawl est une plateforme de SEO technique qui propose des outils d'indexation et d'analyse pour aider les entreprises à prendre des décisions plus éclairées en matière de SEO.
2. **La Search Console de Google**. Conçus par Google, les outils et rapports de la Search Console permettent de mesurer le trafic et les performances d'un site, et de résoudre certains problèmes. Elle peut être utilisée individuellement ou en synergie avec OnCrawl, afin d'obtenir des informations supplémentaires issues de l'analyse combinée de ses propres données et des données d'exploration d'OnCrawl.
3. **L'outil de planification des mots-clés de Google**. Cet outil gratuit permet d'identifier des mots-clés potentiels et le trafic qu'ils génèrent.
4. **L'outil Website Grader de HubSpot**. Il permet à une entreprise d'évaluer le niveau de sécurité de son site et d'optimiser ses performances en apportant quelques modifications à sa stratégie SEO.

En cas de problème spécifique, il peut être utile d'associer différents outils pour vérifier des détails plus pointus, étayer la validité des actions à mettre en place ou, de manière plus générale, mieux comprendre le fonctionnement d'un site.

HubSpot et OnCrawl en recommandent quelques-uns dans la liste suivante, qui, sauf indication contraire, sont gratuits.

Outils génériques :

1. [Les outils de SEO d'OnCrawl](#)
2. [Les outils d'analyse de site de QuickSprout](#)
3. [La barre d'outils de Moz](#)
4. [Le guide sur l'analyse comparative d'AccuRanker](#)

Outils d'indexation :

1. Outil de test de fichiers robots.txt : <https://technicalseo.com/seo-tools/robots-txt/>
2. Outil de création de fichiers sitemap.xml : <https://technicalseo.com/seo-tools/xml-sitemap/>

Outils dédiés aux performances des serveurs :

1. Outil de vérification du code de statut HTTP : <https://next.httpstatus.io/>
2. Outil de test de vitesse d'un site : <https://tools.pingdom.com/>

Outil dédié à la structure d'un site :

1. Outil d'analyse de liens entrants : <https://majestic.com/> (payant)

Outils dédiés aux performances des mots-clés :

1. Outil de recherche et de suivi : <https://fr.semrush.com/> (payant)
2. Outil de recherche : https://ads.google.com/intl/fr_fr/home/tools/keyword-planner/
3. Outil de recherche : <https://keywordtool.io/>
4. Outil de recherche : <https://topics.seomonitor.com/query>

Outil d'optimisation on-page :

1. Outil d'analyse de la vitesse d'une page : <https://developers.google.com/speed/pagespeed/insights/?hl=fr>

Outils de recherche de contenu dupliqué :

1. Outil de recherche de contenu dupliqué interne : <http://www.siteliner.com/>
2. Outil de recherche de contenu dupliqué externe : <http://www.copyscape.com/>

Outils dédiés aux problèmes mobiles :

1. Outil de test d'optimisation mobile : <https://search.google.com/test/mobile-friendly?hl=fr>
2. Outil de validation d'AMP : <https://validator.ampproject.org/>

Outils de test et de validation des balises hreflang :

1. <https://technicalseo.com/seo-tools/hreflang/>
2. <https://www.aleydasolis.com/english/international-seo-tools/hreflang-tags-generator/>
3. <http://hreflang.ninja/>

Outils dédiés aux données structurées :

1. Outil de test de données structurées (Google) : <https://search.google.com/structured-data/testing-tool?hl=Fr>
2. Outil de test des résultats enrichis (Google) : <https://search.google.com/test/rich-results?hl=fr>
3. Outil de validation du balisage (Bing) : <https://www.bing.com/toolbox/markup-validator>

L'indexation



L'indexation

Le SEO technique peut aider une entreprise à définir quelles pages de son site Google pourra parcourir et indexer.

L'un des principaux objectifs d'un audit technique consiste à vérifier la bonne indexation des pages. Certaines doivent être indexées, c'est-à-dire apparaître dans les résultats des moteurs de recherche, tandis que d'autres ne doivent pas être partagées publiquement.

Les pages web que les robots des moteurs de recherche peuvent parcourir sont dites **explorables**. Toutefois, pour apparaître dans les résultats de recherche, elles doivent également être **indexables**. Il faut donc les paramétrer de sorte que les moteurs de recherche puissent les indexer, s'assurer qu'elles sont disponibles lorsqu'une requête est envoyée au serveur où elles sont hébergées, et vérifier qu'elles ne redirigent pas les visiteurs vers une autre page.

Une page indexée peut être affichée dans les résultats d'un moteur de recherche si l'algorithme de ce dernier détermine qu'elle apporte une réponse suffisamment pertinente à une recherche. En d'autres termes, pour apparaître dans les résultats de recherche, une page doit être indexée. Les pages qui ne sont pas indexées ou indexables n'apparaîtront dans aucun résultat de recherche.

Le meilleur moyen de savoir quelles pages d'un site sont indexées par un moteur de recherche consiste à utiliser les informations fournies par ce dernier, comme celles accessibles dans la Search Console de Google ou dans les outils pour webmasters de Bing.

Il est également possible d'obtenir cette information sur Google en saisissant l'URL de la page d'accueil du site avec la commande « site: », par exemple :
site:https://www.exemple.com.

Enfin, des informations supplémentaires sont disponibles en assurant le suivi des pages indexées depuis les fichiers journaux. Pour Google, il suffit par exemple de suivre le nombre de visites de ses robots d'exploration pour bureau et mobile, tout en vérifiant leur identité grâce à leur adresse IP.

Il convient ensuite de comparer les résultats obtenus avec un classement du site établi en fonction du type d'URL.

Un site ne se compose pas uniquement de pages. Le serveur qui l'héberge permet également aux utilisateurs et aux moteurs de recherche d'accéder à d'autres types de fichiers, que l'on appelle généralement des **ressources**, comme :

- Des fichiers .pdf
- Des contenus multimédias (images, vidéos, etc.)
- Des ressources nécessaires au chargement d'une page (fichiers JavaScript, .css ou de polices)

Dans certains cas de figure, il peut être utile qu'un robot d'exploration indexe une ressource précise, comme une image ou un fichier JavaScript qui permet d'ajouter du contenu sur une page. En revanche, il n'est pas nécessaire d'indexer les fichiers de polices, les images par défaut fournies par WordPress ou les fichiers .css de modèles non utilisés.

À indexer :

- Les pages de destination
- Les pages les plus rentables
- Les pages d'accueil
- Les fichiers JavaScript nécessaires à l'affichage d'une page
- Les images destinées à apparaître dans les recherches d'images
- Les pages traduites
- Les pages de promotions saisonnières ou temporaires
- Les actualités
- Les nouvelles pages
- Les pages vers lesquelles d'anciennes pages ont été redirigées

À ne pas indexer :

- Les pages avec du contenu dupliqué qui ne sont pas la principale page à rendre accessible aux utilisateurs
- Les pages d'administration du système de gestion de contenu
- Les aperçus de brouillons de pages
- Les fichiers qui doivent être téléchargés uniquement suite à la soumission d'un formulaire
- Les URL de pages proposant le même contenu, simplement triées ou filtrées à l'aide de requêtes
- Les paginations ou vues de calendrier vierges
- Les pages de confirmation d'inscription, de téléchargement, d'achat, etc.
- Les pages qui ne sont plus disponibles
- Les pages qui ont été redirigées

Pour distinguer les pages indexables d'un site de celles qui ne doivent pas être accessibles aux robots d'exploration, il est également conseillé d'auditer les éléments suivants.

Les fichiers sitemap

Le fichier sitemap.xml d'un site permet d'établir une liste complète des pages qui doivent être accessibles aux moteurs de recherche. Ceux-ci s'attacheront à visiter toutes les URL qui y sont répertoriées, mais ne pas y inscrire une page ne signifie pas qu'elle ne sera pas découverte par Google, notamment si elle est ciblée par les liens d'autres pages.

Le fichier sitemap doit respecter les formats du protocole sitemap disponibles à l'adresse suivante : <https://www.sitemaps.org/fr/protocol.html>. Par ailleurs, il est recommandé de le publier à la racine du site et non pas dans l'un de ses dossiers.

Dans le cadre d'un audit technique, il est impératif de vérifier que le fichier contient uniquement les URL devant être indexées. Enfin, il doit être mis à jour et soumis aux moteurs de recherche régulièrement.

Les fichiers robots.txt

Un fichier robots.txt contient des instructions destinées aux robots d'exploration et représente le premier élément qu'ils visitent lorsqu'ils parcourent un site. Les règles qui y figurent peuvent s'étendre à l'ensemble des robots qui explorent le site ou n'en concerner que quelques-uns.

Un fichier robots.txt s'applique uniquement au domaine ou sous-domaine où il est hébergé et au contenu enregistré dans les fichiers qui y sont stockés. Il n'a pas de valeur pour d'autres domaines ou sous-domaines, même s'il se trouve sur le même serveur qu'eux.

Voici un exemple de règles de fichier robots.txt :

```
user-agent: googlebot-news
disallow: /

user-agent: *
disallow: /wp-admin

user-agent: googlebot
allow: /blog
```

Chaque règle indique le robot d'exploration, ou type de visiteurs, auquel elle s'applique et à qui elle donne ou refuse l'accès à un groupe de fichiers, de pages ou de répertoires.

Questions à se poser avant d'analyser un fichier robots.txt :

- Les robots d'exploration des moteurs de recherche voient-ils les mêmes pages que les visiteurs ?
- Peuvent-ils accéder aux pages qui devraient être indexées ?
- Leur est-il interdit d'accéder aux pages qui ne devraient pas être indexées ?

La valeur des balises méta pour les robots

Il est possible de donner des directives aux robots des moteurs de recherche dans la section <head> de chaque page d'un site.

À l'instar de celles d'un fichier robots.txt, ces directives peuvent être générales comme dans cet exemple,

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

ou spécifiques à un robot précis comme dans celui-ci,

```
<meta name="googlebot" content="noindex" />
```

La liste ci-dessous contient des exemples de balises et leur valeur :

all	N'appliquer aucune restriction. Il s'agit de la valeur par défaut.
noindex	Ne pas indexer cette page. Elle ne fera pas partie de la liste de pages connues de Google et n'apparaîtra pas sur les pages de résultats des moteurs de recherche.
nofollow	Ne suivre aucun lien sur cette page.
none	Identique à noindex et nofollow.
noarchive	Ne pas enregistrer de version archivée disponible pour les utilisateurs de Google.
nosnippet	Ne pas afficher d'aperçu (extrait d'une page ou d'une vidéo) sur les pages de résultats des moteurs de recherche.
notranslate	Ne pas traduire cette page sur les pages de résultats des moteurs de recherche.
noimageindex	Ne pas indexer les images de cette page.

Questions à se poser lors de l'élaboration d'une stratégie dédiée aux balises méta pour les robots :

- Les balises méta utilisées pour les robots permettent-elles d'indexer les pages qui devraient l'être ?
- Les balises méta qui stipulent aux robots de ne pas indexer certaines pages désignent-elles réellement les pages à ne pas indexer ?

Les éléments `rel="alternate"` / `rel="canonical"`

Dans la section `<head>` de chaque page d'un site, il est possible d'utiliser les éléments `<link rel="alternate">` et `<link rel="canonical">` pour indiquer aux moteurs de recherche que plusieurs versions d'une page existent et lui préciser laquelle devrait être indexée.

Ces éléments peuvent être utiles pour un site qui propose des versions mobile et bureau d'une même page. Dans ce cas, il convient d'ajouter les éléments `<link rel="canonical">` à la version mobile et `<link rel="alternate">` à la version bureau pour qu'elles ne soient pas considérées comme redondantes ou en compétition, parce qu'elles cherchent à être référencées pour les mêmes mots-clés.

Les redirections

Les redirections permettent également de stipuler aux robots des moteurs de recherche de ne pas indexer certaines pages d'un site. Elles peuvent être configurées sur le serveur du site et déclencher l'envoi de codes http indiquant que les pages recherchées ont été remplacées de façon définitive (301) ou temporaire (302), ou qu'elles ne sont plus disponibles (410).

Les codes http, notamment les redirections, seront abordés plus en détail dans le chapitre suivant.

Les demandes d'indexation

La plupart des moteurs de recherche permettent aux entreprises de leur envoyer les URL qu'elles souhaitent indexer. Cette fonctionnalité, encore récemment disponible pour tous les utilisateurs, tend aujourd'hui à devenir l'apanage des webmasters et se limite aux URL de leurs propres sites. En septembre 2018, Bing et Google ont supprimé l'accès public à cette option sur leur plateforme respective.

Depuis, Google invite les webmasters à soumettre des demandes d'indexation individuelles pour une URL précise, ou au moyen d'un fichier `sitemap.xml` pour plusieurs URL.

La résolution des problèmes

Il est conseillé de dresser une liste des problèmes d'indexation décelés pour leur apporter des solutions.

À cet effet, le tableau récapitulatif ci-dessous peut être utilisé comme check-list :

Obtenir l'indexation de pages :

- Créer un fichier sitemap
- Autoriser les modifications dans le fichier robots.txt
- Demander l'indexation auprès des moteurs de recherche

Refuser l'indexation :

- Ne pas autoriser les modifications dans le fichier robots.txt
- Ajouter la valeur noindex dans la balise méta pour les robots
- Ajouter l'élément rel="alternate" dans la section <head>
- Configurer des redirections 301, 302 et 410

Sources :

Vérification du robot d'exploration de Google : <https://support.google.com/webmasters/answer/80553?hl=fr>

Fichiers robots.txt : https://developers.google.com/search/reference/robots_txt?hl=fr

Balise méta pour les robots : https://developers.google.com/search/reference/robots_meta_tag?hl=fr

Sitemaps : <https://www.sitemaps.org/fr/>

Indexation d'un site : https://support.google.com/webmasters/answer/34397?hl=fr&ref_topic=3309469

Envoi d'URL : <https://blogs.bing.com/webmaster/september-2018/Anonymous-URL-Submission-Tool-Being-Retired>

Envoi d'URL : <https://developers.google.com/search/docs/guides/submit-URLs?hl=fr>

Les performances des serveurs



Les performances des serveurs

Les performances du serveur qui héberge un site comptent parmi les principaux facteurs qui influencent son classement pour deux raisons majeures. Premièrement, elles participent à la fréquence d'exploration du site, ainsi qu'à son autorité. Deuxièmement, les sites que les utilisateurs considèrent comme lents ou qui présentent trop de dysfonctionnements génèrent moins de clics et favorisent le pogo-sticking, soit le fait pour un utilisateur de revenir sur la page des résultats de sa recherche en cliquant sur le bouton Retour de son navigateur, afin de consulter une autre page.

Un nombre réduit de clics et le pogo-sticking peuvent entraîner une diminution des visites sur un site, mais aussi un déclassement de ses pages dans les résultats de recherche. En effet, l'algorithme de Google tient aussi compte de ce type de comportements chez les utilisateurs pour indexer les pages qu'il parcourt.

Auditer les performances d'un serveur permet de vérifier que :

- Le serveur est opérationnel (aucune erreur de niveau 5xx).
- Le serveur utilise les protocoles de sécurité appropriés (HTTPS).
- Le serveur répond rapidement aux requêtes (vitesse des pages).
- Le serveur permet d'accéder correctement au contenu des pages (vitesse et statut des pages).
- Les erreurs côté serveur nuisent le moins possible à l'expérience utilisateur (pages d'erreur 404 personnalisées).
- Le serveur permet aux robots d'exploration et aux visiteurs d'accéder aux mêmes contenus, comme l'exige Google.

Un robot d'exploration performant doit fournir toutes les informations nécessaires à l'audit d'un serveur. Toutefois, pour comparer les résultats proposés aux robots à ceux présentés aux visiteurs, il est conseillé d'utiliser les données des fichiers journaux du serveur. Certains robots d'exploration conçus pour les audits SEO, comme ceux d'OnCrawl, proposent cette option par défaut.

Le protocole HTTPS

Le protocole HTTPS permet l'envoi d'informations sur internet, comme des requêtes, des URL ou du contenu, et il garantit la sécurité des données échangées entre un visiteur et un serveur.

Pour ce faire, il chiffre les données transférées, vérifie leur intégrité et authentifie l'identité du serveur. Ce protocole de communication est devenu la norme et les protocoles moins sécurisés, comme le HTTP, peuvent nuire au référencement d'un site.

L'audit de son serveur doit donc permettre à une entreprise de déterminer si son site applique les normes et protocoles appropriés. Les sites n'ayant pas encore opéré la migration vers le protocole HTTPS peuvent utiliser ce [guide](#) de Google.

La vitesse d'un site

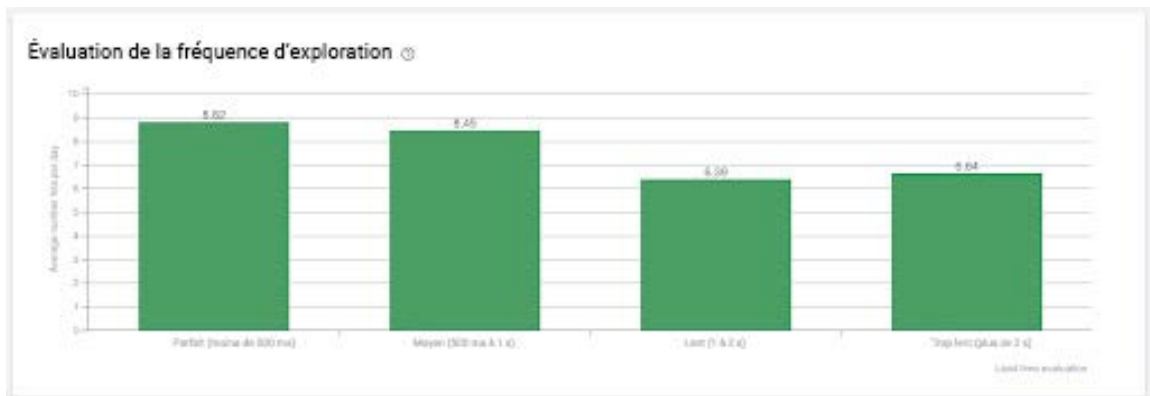
La vitesse d'un site web compte parmi les facteurs de référencement des moteurs de recherche, en particulier pour les recherches effectuées sur appareils mobiles. Elle permet de départager deux sites identiques dans les résultats de recherche, en privilégiant le plus rapide.

Évaluer la vitesse des pages d'un site dans le cadre d'un audit technique sert un triple objectif : déterminer le seuil de vitesse à partir duquel les visites des robots et celles issues de la recherche naturelle diminuent, établir une liste exhaustive des pages dont la vitesse de chargement se situe en deçà de ce seuil, et identifier les causes du problème.

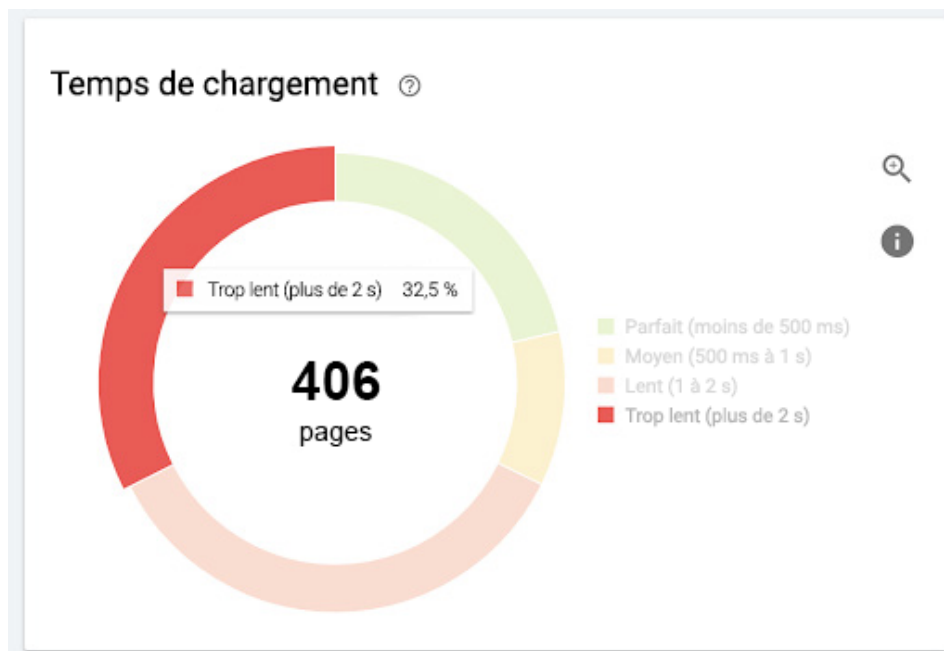
Pour atteindre ce triple objectif, il est recommandé aux entreprises de combiner l'analyse de leurs fichiers journaux et les résultats de l'audit du SEO technique de leur site.



Plus une page est lente (rouge), moins elle est susceptible d'attirer des visites naturelles (barres sur la gauche). Source : OnCrawl et données des fichiers journaux.



De manière générale, plus une page charge rapidement, plus elle est parcourue fréquemment par les robots d'exploration. Ici, améliorer le temps de chargement des pages qui prennent plus de 1 seconde à s'ouvrir semble plus pertinent que de se concentrer sur celles dont la vitesse de chargement est déjà inférieure à 1 seconde. Source : OnCrawl et données des fichiers journaux.



406 pages ADD COLUMNS

FULL URL	DEPTH	LOAD TIME (IN MS)	LOAD TIME EVALUATION
https://www.oncrawl.com/...	2	2.520	Too slow (> 2s)
https://www.oncrawl.com/...	2	2.567	Too slow (> 2s)
https://www.oncrawl.com/...	2	2.609	Too slow (> 2s)
https://www.oncrawl.com/...	2	4.565	Too slow (> 2s)
https://www.oncrawl.com/...	2	3.789	Too slow (> 2s)
https://www.oncrawl.com/...	2	2.690	Too slow (> 2s)

Accès à une liste de pages considérées comme étant trop lentes par un robot d'exploration. Source : OnCrawl.

Lorsque le serveur répond correctement, le temps de chargement d'une page dépend essentiellement de sa taille, qui comprend des éléments comme sa longueur et les ressources nécessaires à l'affichage de son contenu. Ces ressources incluent les images, les vidéos, les fichiers intégrés .pdf, .css et JavaScript ainsi que les polices, qui doivent être chargés pour permettre l'affichage de la page.

Une fois les pages les plus volumineuses identifiées, il est possible d'en réduire la taille en suivant les conseils ci-dessous :

Problèmes liés au contenu :

- Compresser le HTML
- Minifier les fichiers JavaScript et .css
- Réduire la taille des images
- Déplacer les scripts qui nuisent à l'affichage de la page pour qu'ils s'exécutent à une étape ultérieure du processus de chargement, si possible

Mise en cache :

- Utiliser le cache du navigateur
- Utiliser un CDN (réseau de diffusion de contenu)

Réponses du serveur :

- Améliorer la réactivité du serveur
- Réduire le nombre de redirections

Le statut d'une page

Toute requête relative à une URL envoyée à un serveur reçoit en réponse un code de statut indiquant la disponibilité du contenu associé à cette URL. Par exemple, les codes d'erreur 404 (document introuvable) ou 503 (serveur indisponible) indiquent aux moteurs de recherche que les informations d'un site ne sont pas toujours disponibles pour être utilisées comme résultats de recherche. Le classement d'une page peut donc être affecté si elle renvoie trop souvent des messages d'erreur (codes de niveau 4xx et 5xx).

À l'instar du code de statut 200 qui indique le succès d'une requête, certains codes sont intentionnels, comme les redirections 301 qui permettent, par exemple, de rediriger les utilisateurs vers la nouvelle URL d'une page lorsque l'ancienne a été modifiée. Les messages de redirection (code de niveau 3xx) fournissent des informations essentielles aux moteurs de recherche lorsqu'un site subit des modifications. En outre, Google a clairement indiqué que les codes de niveau 3xx améliorent l'autorité et le classement des URL vers lesquelles ils redirigent, et qui sont vouées à remplacer les anciennes URL dans les pages de résultats.

Les codes de statut les plus courants sont :

Codes de niveau 2xx : les messages de succès

200 OK

Codes de niveau 3xx : les messages de redirection

301 Document déplacé définitivement
302 Document déplacé temporairement
304 Document inchangé depuis la dernière requête
307 Redirection temporaire

Codes de niveau 4xx : les messages d'erreur côté client (navigateur)

400 Mauvaise requête
401 Accès interdit sans identification
404 Document introuvable
410 Ressource indisponible ou supprimée

Codes de niveau 5xx : les messages d'erreur côté serveur (site web)

500 Erreur interne du serveur
503 Service indisponible
550 Permission refusée

Un audit technique doit permettre d'identifier :

- Toutes les pages déclenchant des messages d'erreur côté serveur (codes de niveau 5xx). Ces erreurs doivent être rectifiées.
- Toutes les pages qui renvoient des messages d'erreur 4xx. Bien qu'elles n'affectent pas nécessairement le SEO d'un site, elles sont source de frustration pour les utilisateurs. Elles doivent être mises à jour ou redirigées au moyen d'un code de niveau 3xx.
- Toutes les pages redirigées avec un code de niveau 3xx et les pages vers lesquelles elles redirigent.

Les chaînes et boucles de redirections

Un site faisant l'objet de modifications en série peut générer des chaînes de redirections où une page A est redirigée vers une page B, qui est redirigée vers une page C, et ainsi de suite.

Les robots de Google cessent d'explorer les redirections au bout de 5 liens. Au bout d'une vingtaine de liens, les navigateurs affichent un message d'erreur à l'attention des visiteurs. Ce seuil varie selon les navigateurs.

Il arrive également que certaines chaînes de redirections se referment sur elles-mêmes pour former des boucles : la page C n'est pas redirigée vers la page D, mais vers la B, ce qui crée une boucle sans fin.

Les robots d'exploration qu'une entreprise utilise sur son propre site doivent donc tenir compte des pages avec des codes de niveau 3xx pour lui permettre de déceler les chaînes et les boucles de redirections.

Un audit technique doit :

- Identifier toutes les pages qui font partie de chaînes ou de boucles de redirections.
- Identifier toutes les chaînes et toutes les boucles qui contiennent plus de 5 liens.
- Identifier les liens redirigeant vers des pages avec un code de niveau 3xx comprises dans une chaîne de redirections. Ces liens doivent être mis à jour pour rediriger vers la dernière page de la chaîne.

Les pages d'erreur 404 personnalisées et les erreurs de type soft 404

Les erreurs 404 sont impossibles à éradiquer. Même si toutes les pages et tous les liens d'un site sont parfaitement opérationnels, il peut arriver qu'un utilisateur saisisse une URL incorrectement ou qu'il soit impossible d'obtenir la mise à jour d'un lien externe suite à la modification d'une URL. Les pages d'erreur 404 personnalisées permettent de prévenir ces cas de figure en assurant la fluidité de l'exploration des robots et de l'expérience des utilisateurs.

Toutefois, il n'est pas recommandé de rediriger ou de remplacer automatiquement une page obsolète par une page vide personnalisée, d'erreur ou sans rapport, au risque de provoquer des erreurs de type « soft 404 ». Lorsque celles-ci se produisent, l'utilisateur accède à ce qui semble être une page d'erreur 404, alors qu'il est indiqué au navigateur et aux robots d'exploration qu'il s'agit d'une URL opérationnelle. Elles représentent donc une perte de temps pour les robots et une source de frustration pour les utilisateurs.

L'un des objectifs d'un audit technique consiste à vérifier qu'un site comprend une page d'erreur 404 personnalisée opérationnelle, contenant au moins un lien permettant aux visiteurs de rester sur le site.

Sources :

Protocole HTTPS : <https://support.google.com/webmasters/answer/6073543?hl=fr>

Vitesse d'une page (2018) : <https://webmaster-fr.googleblog.com/2018/01/vitesse-chargement-pages-critere-positionnement.html>

Vitesse d'une page (générique) : <https://fr.oncrawl.com/referencement/vitesse-des-pages-impact-seo/>

Indicateurs de référence en matière de vitesse (2017) : <https://think.storage.googleapis.com/docs/mobile-page-speed-new-industry-benchmarks.pdf>

La structure d'un site



La structure d'un site

La structure d'un site représente l'une des pierres angulaires d'un audit technique, car elle revêt une importance capitale pour les robots d'exploration de Google.

Elle comprend la composition des URL, le type de connexions établies entre les pages du site, et les éléments qui permettent aux utilisateurs et aux moteurs de recherche de comprendre l'objectif du site, les sujets traités et les catégories utilisées.

blog.hubspot.fr/marketing/recherche-mots-cles-seo

De nombreuses sources, notamment la page des consignes aux webmasters disponible dans le centre d'aide de Google, indiquent que l'architecture d'un site compte parmi les éléments analysés par les robots d'exploration. L'article de support de la plateforme préconise d'ailleurs de créer une arborescence de pages claire et conceptuelle.

Les robots d'exploration de Google définissent les caractéristiques d'un site à partir de ce qu'ils comprennent de son organisation. Ils peuvent par exemple utiliser sa structure pour proposer des résultats sous forme de catégories thématiques, comme dans l'exemple ci-dessous.

<https://europa.eu> > european-union > index_fr ▼

Site web officiel de l'Union européenne | Union ... - EUROPA

17 mai 2019 - European Union - Official website of the European Union.

European Union (EU)

About the EU - EU law -
Countries - Brexit - EU Newsroom -
...

About the EU

The EU in brief - Countries -
Institutions and bodies - ...

À propos de l'UE

Comment l'UE est financée - À
quoi le budget de l'UE est ...

Pays

Espagne - France - Italie - Croatie -
...

Domaines d'action de l'Union ...

Domaines d'action de l'Union
européenne. EU Login ...

Countries

EU countries, candidate countries
and other European countries.

Facile à lire - L'Union ...

L'Union européenne est un groupe
de 28 pays en Europe. Ces ...

Press releases

Communiqués de presse – toutes
les bases de données. Latest ...

Espace Apprentissage

Espace Apprentissage - Matériel
pédagogique, jeux et bien plus ...

L'UE en bref

Les valeurs de l'Union européenne
sont communes ...

Les pages connectées les unes aux autres sur un même site sont source de notoriété et d'autorité. Une structure de liens internes bien définie permet aux robots de Google d'identifier les pages les plus importantes. Les stratégies qui s'articulent autour de la structure des liens internes d'un site participent à l'optimisation du classement de certaines de ses pages. Elles sont donc souvent utilisées dans le cadre de la promotion de pages spécifiques, au cours de périodes limitées dans le temps.

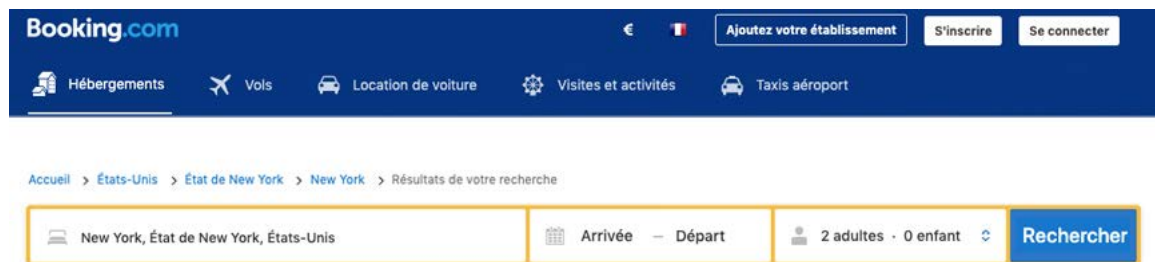
Un audit technique doit prendre en compte différents éléments relatifs à l'organisation d'un site pour définir sa structure et identifier les leviers d'amélioration.

Les fils d'Ariane

Ils correspondent à une liste de catégories de plus en plus précises qui retrace le chemin d'accès de la page visitée et se trouve généralement sous son titre et au-dessus de son contenu.

Dans de nombreux cas, ils s'apparentent à un type de structure navigationnelle, même s'il n'est pas nécessaire qu'ils soient fidèles au menu de navigation.

Ils peuvent également être utilisés pour établir une hiérarchie d'attributs retraçant le chemin d'accès jusqu'à la page visitée, comme sur le site de Booking.com : les pages d'hébergements sont classées par destinations, bien qu'il soit difficile de suivre un chemin de navigation de la page d'accueil aux hébergements situés aux États-Unis, puis dans l'État de New York et enfin dans la ville de New York.



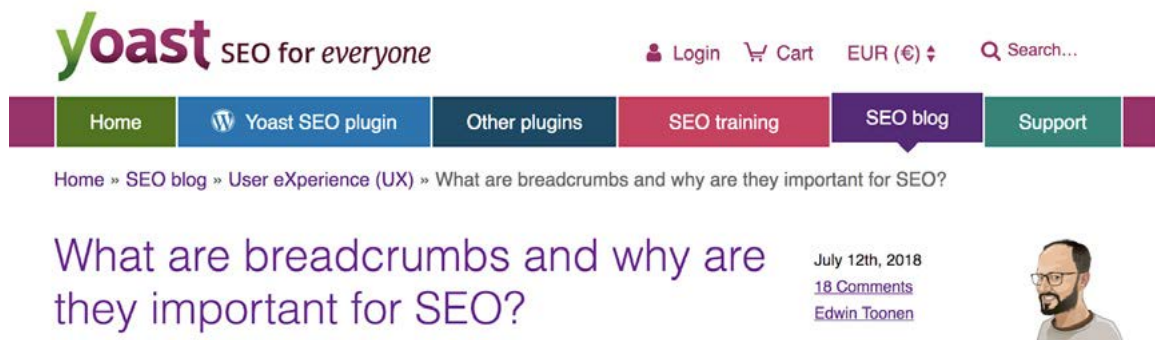
Ce type de fil d'Ariane est courant sur les sites d'e-commerce, où les attributs correspondent souvent à des catégories de produits, comme ci-dessous sur le site amazon.fr.



Les fils d'Ariane jouent un rôle important dans le SEO, car en présentant les principales catégories structurelles ou catégories d'attributs d'un site, ils aident les moteurs de recherche à en interpréter le contenu. En outre, ils sont organisés de façon structurée et donc facilement accessibles et compréhensibles pour les robots d'exploration.

Ils favorisent une expérience utilisateur de meilleure qualité puisqu'ils permettent aux visiteurs de poursuivre leur visite plus facilement, et aux moteurs de recherche de s'assurer qu'un site apporte des réponses pertinentes aux questions des utilisateurs. Ceci est particulièrement vrai pour les utilisateurs d'appareils mobiles, sur lesquels les fils d'Ariane sont souvent plus faciles à voir et à suivre qu'un menu de navigation principal exhaustif.

Il est possible d'utiliser des données structurées pour générer un fil d'Ariane et ainsi permettre à Google de l'afficher dans l'extrait descriptif d'une page, comme dans l'exemple du fil d'Ariane de navigation de l'article de blog ci-dessous :



The screenshot shows the top navigation bar of the Yoast website. The logo 'yoast SEO for everyone' is on the left. On the right, there are links for 'Login', 'Cart', 'EUR (€)', and a search bar. Below the navigation bar is a horizontal menu with categories: 'Home', 'Yoast SEO plugin', 'Other plugins', 'SEO training', 'SEO blog' (highlighted), and 'Support'. Below the menu is a breadcrumb trail: 'Home » SEO blog » User eXperience (UX) » What are breadcrumbs and why are they important for SEO?'. The main heading of the article is 'What are breadcrumbs and why are they important for SEO?'. To the right of the heading, the article metadata is displayed: 'July 12th, 2018', '18 Comments', and 'Edwin Toonen' with a small profile picture of a man with glasses and a beard.

Les données structurées sont abordées plus en détail dans le chapitre intitulé Schema.org.

Les fils d'Ariane sont particulièrement utiles pour les types de sites suivants :

- Sites d'e-commerce
- Blogs volumineux
- Bases de connaissances
- Sites web vastes ou complexes
- Applications web

Un audit technique doit permettre de définir l'emplacement des fils d'Ariane sur un site en apportant des réponses aux questions suivantes :

- Le site utilise-t-il des fils d'Ariane ?
- Les fils d'Ariane sont-ils utilisés de façon systématique ?
- Les noms des catégories des fils d'Ariane sont-ils clairs et faciles à comprendre ?
- Les fils d'Ariane sont-ils utilisés aux bons endroits ?

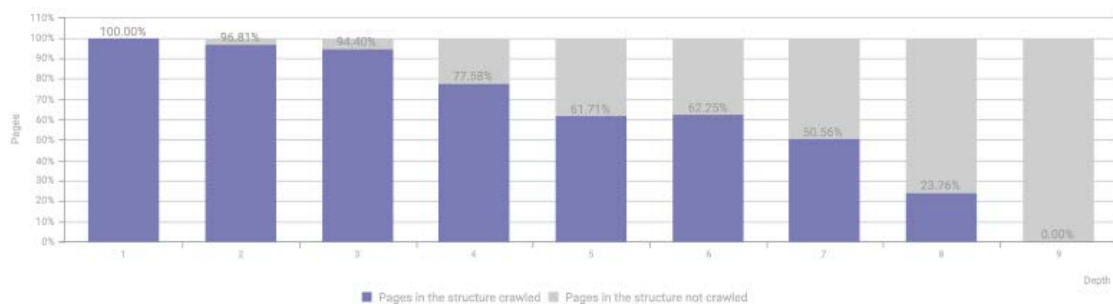
La profondeur de page

La profondeur de page fait référence au nombre de liens sur lesquels un utilisateur doit cliquer avant d'atteindre la page souhaitée. Il s'agit là encore d'un facteur qui a des répercussions sur le SEO d'un site, à la fois du point de vue des utilisateurs et de celui des moteurs de recherche.

Un aperçu des comportements des utilisateurs est disponible dans Google Analytics, depuis la rubrique Flux de comportement du menu Comportement. Il indique les taux de clics de chaque lien par lequel passe un utilisateur avant d'accéder à la page qu'il recherche. Plus le nombre de liens est élevé, moins l'utilisateur est susceptible d'accéder à la page souhaitée.

Le même principe s'applique aux robots d'exploration. Suite à l'analyse de plusieurs centaines de sites web, OnCrawl a constaté que plus une page est éloignée de la page d'accueil, moins Google a tendance à l'explorer. Ce constat a depuis été confirmé par Google, qui considère en effet cette distance comme un critère de classement.

Pages crawled and not crawled by depth ⓘ



Un audit technique doit permettre de déterminer :

- Le nombre et le type de pages d'un site en fonction de leur profondeur.
- L'impact de la profondeur d'une page sur son indexation et son classement, en comparant le nombre de pages explorées au nombre de pages indexées en fonction de leur profondeur.

La stratégie dédiée aux liens internes

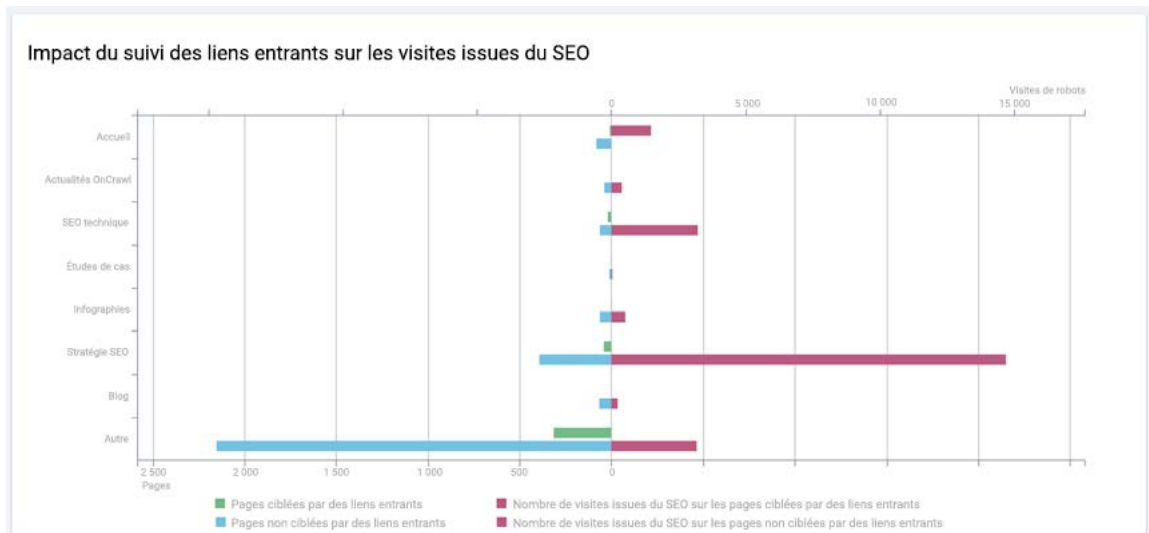
La stratégie dédiée aux liens internes d'un site entre en action dès lors qu'un audit permet d'identifier des points d'amélioration.

Les liens entrants

Il s'agit des liens de sites externes qui redirigent vers les URL internes du site d'une entreprise donnée. Les moteurs de recherche les utilisent pour établir l'autorité d'un site sur un sujet particulier. Ce facteur, important pour le classement des sites pour des mots-clés en concurrence les uns avec les autres, explique pourquoi la fiabilité et la pertinence des liens entrants participent à la réussite d'une stratégie SEO.

Ces liens permettent en outre de booster le trafic d'un site. Ils doivent ouvrir des pages intégrées à la structure du site vers lequel ils redirigent par le biais de liens internes, et doivent également conférer davantage d'autorité à d'autres pages traitant de sujets similaires. Les pages accessibles depuis des liens entrants doivent être parfaitement opérationnelles pour faire une bonne première impression.

En déterminant l'impact des liens entrants sur le trafic naturel et sur le comportement des robots d'exploration, ainsi que leur distribution en fonction d'informations comme le type ou la profondeur de page, l'analyse croisée des données d'exploration et des liens entrants représente l'un des principaux facteurs de réussite d'un audit technique.



Association d'OnCrawl et de Majestic pour établir une corrélation entre les liens entrants et le trafic naturel.

Il est important de noter que les liens entrants provenant de sites peu fiables, dangereux ou traitant de sujets totalement différents peuvent nuire à la stratégie SEO d'une entreprise.

Un audit technique doit permettre de répondre aux questions suivantes :

- Quelles pages sont accessibles depuis un lien entrant ?
- Les pages accessibles depuis un lien entrant sont-elles opérationnelles ?
 - Sont-elles également accessibles depuis des liens internes ?
 - Le code de statut 200 est-il bien en place ?
- Quels sites hébergent les liens entrants ?
- Existe-t-il une corrélation entre le fait qu'une page soit ciblée par des liens entrants et ses performances ?
- L'amélioration des performances d'une page accessible depuis des liens entrants se propage-t-elle aux autres pages du site reliées par des liens internes ?

Les liens internes et la notoriété

Les liens internes sont les outils qui permettent de créer l'architecture d'un site et de transférer la notoriété d'une page à une autre.

Les bonnes pratiques en matière de liens internes sont les suivantes :

- Limiter le nombre de liens par page, quelques milliers tout au plus, selon les consignes aux webmasters de Google. Il est important de noter que ce total inclut tous les liens dans les titres, menus, menus latéraux et pieds de page.
- Ajouter des liens sur les pages principales, comme la page d'accueil, vers les pages de catégories, puis sur les pages de catégories vers toutes les pages d'une même catégorie, afin de regrouper le contenu par type, thème ou catégorie.
- Créer des liens entre les différentes catégories.
- Ajouter des liens sur les pages performantes pour qu'elles confèrent leur notoriété à d'autres pages.
- Ajouter des liens vers les pages à promouvoir.

Un audit technique doit permettre de répondre aux questions suivantes :

- Vers quel groupe de pages les liens d'un autre groupe de pages redirigent-ils ?
- Quel groupe compte le plus de pages ciblées par des liens ?
- Existe-t-il une relation de cause à effet entre l'utilisation de liens internes et le nombre de visites des robots d'exploration ? Utiliser le profil optimal pour améliorer les performances des pages qui n'atteignent pas leurs objectifs.
- Existe-t-il une corrélation entre lien interne et visite naturelle ? Utiliser le profil optimal pour améliorer les performances des pages qui n'atteignent pas leurs objectifs.
- Le site contient-il des pages ciblées par un trop grand nombre de liens ?
- Le site contient-il des pages n'étant ciblées par aucun lien ?

Les liens erronés

Il s'agit de liens qui ne permettent pas à un utilisateur d'accéder à la page qu'il recherche. Ils sont souvent le résultat de pages contenant un code de statut de niveau 4xx (messages d'erreur côté client) ou 5xx (messages d'erreur côté serveur).

Outre leur impact sur l'expérience utilisateur, les moteurs de recherche rétrogradent voire désindexent les pages introuvables, car elles n'apportent aucune réponse aux recherches des utilisateurs.

Certains moteurs de recherche peuvent considérer les codes de niveau 3xx comme des liens erronés. Bien que les liens de redirection permettent à un utilisateur d'accéder à la page souhaitée, ils n'opèrent pas la redirection eux-mêmes. Par ailleurs, Google n'indexera pas une page que ses robots ont identifiée comme la cible d'une redirection 3xx. Il est donc important que des liens internes redirigent vers la page mise à jour avec un code de statut 200 afin qu'elle puisse être indexée.

Un audit technique doit permettre de créer une liste de liens à mettre à jour par catégorie :

- Liens internes vers une page d'erreur 4xx ou 5xx.
- Liens internes vers une page de redirection 3xx.

Les sous-domaines

Les robots d'exploration de Google traitent les sous-domaines (<https://sousdomaine.exemple.com/page/>) et les sous-dossiers (<https://www.exemple.com/sousdossier/page/>) de la même façon.

De nombreuses stratégies SEO favorisent les sous-dossiers, car hormis dans le cadre de l'exploration, Google les considère comme des entités à part entière. Les moteurs de recherche doivent fournir davantage d'efforts pour maintenir et promouvoir les classements de sous-domaines qui peuvent être traités comme des sites distincts. Ils les référencent donc séparément et la notoriété des uns n'est pas transférée aux autres.

D'autres stratégies sont axées sur les sous-domaines, en particulier lorsqu'elles s'appliquent à des types de contenus très variés.

Sauf contre-indication, un audit technique doit traiter les sous-domaines comme faisant partie intégrante du site audité et se concentrer sur les mêmes sujets pour les sous-domaines et le site principal.

Les pages orphelines

Il s'agit de pages qui ne sont pas incluses dans la structure d'un site. Elles ne possèdent pas de page parent et ne sont ciblées par aucun lien interne.

Les pages orphelines ne profitent d'aucun des attributs susceptibles de les mettre en valeur, tels que des liens permettant d'y accéder depuis des pages importantes ou des pages traitant de sujets similaires, ou encore un emplacement stratégique à quelques clics seulement de la page d'accueil. Par conséquent elles ne sont jamais aussi bien classées qu'elles pourraient l'être.

En revanche, elles peuvent produire des résultats rapides, car il suffit de les intégrer à la structure du site pour optimiser leurs performances et leur classement.

Elles sont souvent le résultat de la restructuration d'un site, de l'expiration d'une page, ou de la suppression de pages ou de liens hors audit.

Un audit technique doit :

- Identifier toutes les pages orphelines en comparant les pages explorées aux :
 - Pages indexées
 - Pages visitées
 - Pages référencées dans le fichier sitemap
- Répertorier les pages orphelines actives, c'est-à-dire celles qui apparaissent dans les résultats de recherche et génèrent des visites naturelles. Ces pages atteignent déjà une partie de leurs objectifs et produiront rapidement de meilleurs résultats si elles sont optimisées.

Sources :

Consignes aux webmasters : <https://support.google.com/webmasters/answer/35769?hl=fr>

Fils d'Ariane : <https://developers.google.com/search/docs/data-types/breadcrumb?hl=fr>

Profondeur de page : <https://fr.oncrawl.com/seo-technique/distribution-profondeur-page-urls/>

Liens internes : <https://blog.hubspot.fr/marketing/seo-groupes-thematiques>

Sous-domaines : <https://youtu.be/uJGDyAN9g-g>

Pages orphelines : <https://fr.oncrawl.com/pages-orphelines/>

Les performances des mots-clés



Les performances des mots-clés

Comprendre les performances des mots-clés constitue l'un des principaux enjeux d'un audit technique. Les mots-clés sont évalués essentiellement sur la base du volume de trafic naturel qu'ils génèrent pour les pages d'un site. Pour définir leurs performances, il est nécessaire d'analyser la stratégie de recherche de mots-clés et les URL classées dans les résultats de recherche du site. Ce travail d'analyse permet d'identifier les caractéristiques techniques des pages les plus performantes, ainsi que celles des pages qui gagneraient à être optimisées. Lorsqu'un problème technique lié, par exemple, à la vitesse de chargement a des répercussions sur l'ensemble des pages devant être mises à jour, un simple ajustement peut produire des résultats impressionnants.

Identifier les problèmes techniques susceptibles d'affecter les performances des mots-clés demande cependant de cibler les bons mots-clés et de suivre leur classement dans les résultats de recherche.

La recherche de mots-clés

Qu'une entreprise soit en plein développement ou déjà bien établie, la recherche de mots-clés constitue l'un des piliers de la réussite de sa stratégie de SEO technique. Les mots-clés évoluent au rythme de l'entreprise et doivent donc être évalués régulièrement, notamment lors d'un audit technique.

La recherche de mots-clés en **6 étapes** :

1. Établir une liste de sujets importants et pertinents pour l'entreprise.
Quels sont les thèmes récurrents dans ses articles de blog ? Quels sujets sont abordés lorsque l'on parle d'elle dans une discussion ? Les réponses à ces questions permettent de définir l'entreprise en des termes génériques.
2. Affiner ces termes pour en faire des mots-clés ciblés et pertinents.
Les expressions ou les termes qui sont importants pour le classement dans les résultats de recherche.
3. Rechercher des mots-clés associés.
Compléter la liste avec des mots-clés similaires et connexes en effectuant une recherche sur Google ou en utilisant d'autres outils de recherche de mots-clés.
4. Définir des mots-clés génériques et des mots-clés de longue traîne.
Les premiers sont moins ciblés et souvent populaires, les seconds sont les versions plus détaillées des premiers. Les mots-clés de longue traîne se présentent généralement sous la forme d'expressions et ils obtiennent de meilleurs classements dans les résultats de recherche, car ils sont moins en concurrence les uns avec les autres.
5. Analyser le classement des entreprises concurrentes pour les mots-clés ciblés.
Se concentrer à la fois sur les mots-clés utilisés par la concurrence mais pas l'entreprise, et ceux utilisés par l'entreprise et ses concurrents. Des outils comme [SEMrush](#) permettent de réaliser des analyses comparatives.
6. Affiner la liste à l'aide de l'[outil de planification des mots-clés de Google](#).
Analyser le volume de trafic estimé pour les mots-clés de la liste afin d'identifier ceux qui génèrent trop ou trop peu de recherches.

Une fois la liste de mots-clés dressée, il est nécessaire d'évaluer les performances actuelles des pages où ils apparaissent, afin d'identifier les points forts et les points faibles du site.

Les URL classées

Il s'agit d'URL qui apparaissent dans les résultats de recherche et témoignent du fait qu'une stratégie SEO a l'effet escompté sur les mots-clés ciblés. Le classement d'une URL correspond à sa position quotidienne sur les pages de résultats des moteurs de recherche.

Lorsqu'une entreprise suit le classement de ses pages web en tenant compte des mots-clés pour lesquels elle souhaite être référencée, elle peut identifier les URL qui réalisent de bonnes performances et celles qui nécessitent d'être optimisées. En se concentrant sur les mots-clés qui n'atteignent pas leurs objectifs, elle pourra améliorer le référencement de son site dans la recherche naturelle.

Les données sur le classement des mots-clés disponibles dans la Search Console de Google permettent également de mieux comprendre l'impact de certains facteurs techniques sur leurs performances.

Un audit technique doit permettre d'établir s'il existe une corrélation entre les performances des mots-clés et des facteurs comme :

- Le type de page, qu'il s'agisse d'une page produit, de blog ou de destination
- La longueur des mots-clés
- Les mots-clés génériques et les mots-clés de marque
- La profondeur d'une page, ou le nombre de clics qui la séparent de la page d'accueil
- Le nombre de liens internes redirigeant vers la page
- Le temps de chargement de la page
- Le nombre de mots
- L'optimisation de la balise de titre
- Les indicateurs de popularité internes, comme le classement Inrank établi par OnCrawl
- La longueur des méta-descriptions
- L'utilisation de données structurées
- La fréquence des explorations

Le trafic naturel

Pour organiser les URL classées d'un site et évaluer les performances de ses mots-clés, il est important de pouvoir déterminer le volume de trafic naturel généré par ces URL. Cette opération est possible directement dans OnCrawl, en utilisant les données sur les visites naturelles disponibles dans les fichiers journaux du site.

Elle peut également être réalisée manuellement, si les URL classées sont analysées dans Google Analytics, en suivant les étapes ci-dessous :

1. Accéder à Google Analytics.
2. Sélectionner Acquisition, Campagnes puis Mots-clés naturels pour afficher le trafic naturel du site.
3. Définir le nombre de mots-clés et la période à analyser (par exemple, 30 jours et 500 lignes).
4. Cliquer sur Dimension secondaire puis sur Mot-clé pour afficher les mots-clés qui apparaissent dans les pages de destination sélectionnées au cours de l'étape précédente.
5. Cliquer sur Exporter pour enregistrer les données dans une feuille de calcul Excel ou Google.
6. Conserver les colonnes Mot-clé, Page de destination, Sessions, Pages/Visite et Taux de rebond, et supprimer les colonnes inutiles.
7. Ajouter les colonnes Sessions potentielles et Performance.
8. Ajouter des filtres aux titres des colonnes pour filtrer les mots-clés par session.
9. Ouvrir l'[outil de planification des mots-clés de Google](#).
10. Copier les mots-clés de la colonne 1 de la feuille de calcul Excel ou Google et les coller dans l'outil de planification dans le même ordre.
11. Dans l'outil, sélectionner le type de correspondance [mot-clé exact].
12. Exporter les résultats en cliquant sur Télécharger, puis sur Tous les résultats de recherche et sur CSV.
13. Importer l'ensemble de données dans la colonne Sessions potentielles de la feuille de calcul.
14. Définir la colonne Performance comme égale à la moyenne des colonnes Visites et Sessions potentielles, puis convertir le résultat en pourcentage.
15. Analyser la colonne Performance pour déterminer le volume de trafic potentiel généré par chaque mot-clé.

Un audit technique doit :

- Indiquer le trafic naturel de chaque URL classée.
- Établir un classement décroissant de ces pages par nombre de visites.

Sources :

Recherche de mots-clés : <https://blog.hubspot.fr/marketing/recherche-mots-cles-seo>

Mots-clés génériques et mots-clés de longue traîne : <https://research.hubspot.com/charts/example-of-head-term-vs-long-tail-keywords>

L'optimisation on-page



L'optimisation on-page

L'optimisation on-page englobe toutes les actions qui peuvent être mises en place pour améliorer les performances SEO d'une page précise. Elle cible les problèmes qui affectent cette page en particulier plutôt que l'ensemble du site.

Les éléments on-page comprennent le contenu de la page, mais aussi les éléments méta de son code source qu'utilisent les moteurs de recherche pour obtenir des informations à son sujet.

Les balises HTML utilisées dans le cadre d'une stratégie SEO

Les balises HTML permettent de structurer le contenu d'une page, et aident les robots d'exploration et les visiteurs à l'interpréter et la comprendre.

Les balises HTML les plus importantes pour le SEO incluent actuellement les balises méta comme la balise de titre, la balise méta-description, la balise méta pour les robots, ou encore les balises d'en-têtes H1, H2, etc.

Les balises de titre

Les balises de titre sont essentielles à la création et à l'audit de contenu on-page. Les titres servaient initialement à libeller les fenêtres et les onglets d'un navigateur. Aujourd'hui, ils apparaissent encore dans le libellé des onglets et constituent donc l'une des rares balises méta visibles par les utilisateurs. Les balises de titre jouent en outre un rôle important dans l'indexation et le classement d'une page.

Les bonnes pratiques à appliquer aux balises de titre sont les suivantes :

- Elles doivent inclure les principaux mots-clés susceptibles d'être saisis dans les moteurs de recherche par l'audience ciblée. Même lorsqu'ils se situent après la limite de caractères, ces mots-clés participent au classement de la page.
- Les titres sont tronqués différemment selon le type d'appareil sur lesquels ils apparaissent. Il est donc conseillé de respecter l'ancienne limite de 60 caractères pour qu'ils s'affichent correctement, quel que soit l'appareil utilisé.
- Privilégier les mots-clés de longue traîne pour faciliter l'optimisation des performances du site.

Un audit technique doit permettre de :

- S'assurer que chaque page possède un titre unique.
- S'assurer que les 60 premiers caractères du titre d'une page révèlent ce dont elle traite.

Les méta-descriptions

La description ou l'aperçu d'une page qui s'affiche dans les résultats de recherche provient près d'une fois sur trois de la méta-description qui sert à définir brièvement le contenu qu'elle propose.

Google s'attache à adapter la description d'une page aux requêtes formulées par les utilisateurs. Il est donc possible que la description change en fonction des termes de recherche. Si Google ne peut pas utiliser la méta-description, l'algorithme sélectionnera un extrait pertinent du contenu de la page.

Les bonnes pratiques à appliquer aux méta-descriptions sont les suivantes :

- Insérer une balise de description unique et personnalisée sur chaque page du site pour éviter que les moteurs de recherche n'en sélectionnent eux-mêmes les premières phrases comme extrait descriptif.
- Rédiger une méta-description de 70 à 155 caractères. Il est important de noter que cette fourchette varie au fil de l'évolution des pratiques de Google. Sa valeur maximale était, par exemple, fixée à 320 caractères durant les premiers mois de l'année 2018.
- Ne pas utiliser de guillemets, car Google les supprime. Le cas échéant, préférer les guillemets simples (') aux doubles (« »).
- Ne pas mentionner les mots-clés de façon excessive.
- Rédiger des textes uniques qui évitent les lieux communs. Dans le cas contraire, Google risque de ne pas afficher de description.
- Inclure des informations sur la page en séparant clairement chaque terme, quitte à altérer la syntaxe de la phrase.

Un audit technique doit :

- Vérifier que toutes les pages du site ont une méta-description.
- Identifier les méta-descriptions redondantes ou génériques.
- Identifier les méta-descriptions trop courtes ou trop longues.

Les balises méta pour les robots

Les balises méta pour les robots permettent de donner des instructions d'exploration et d'indexation aux robots des moteurs de recherche quant au comportement à adopter sur les pages qu'ils parcourent. Il en existe deux types.

- Les balises « **index/noindex** » indiquent aux moteurs de recherche s'ils doivent ou non afficher une page dans les résultats de recherche.
- Les balises « **follow/nofollow** » indiquent aux moteurs de recherche s'ils doivent tenir compte des liens et s'ils doivent les suivre ou non d'une page vers la suivante.

Il est possible d'associer ces deux balises de différentes façons :

- « index, follow » : les robots indexeront la page parcourue et visiteront les pages vers lesquelles elle redirige.
- « index, nofollow » : les robots indexeront la page parcourue mais ne visiteront pas les pages vers lesquelles elle redirige.
- « noindex, follow » : les robots n'indexeront pas la page parcourue mais visiteront les pages vers lesquelles elle redirige.
- « noindex, nofollow » : les robots n'indexeront pas la page parcourue et ne visiteront pas les pages vers lesquelles elle redirige.

À ce stade, l'audit technique ayant déjà couvert les balises méta pour les robots dans le cadre de l'indexabilité des pages, il devra s'assurer que les balises de page sont utilisées à bon escient.

Les en-têtes

Les en-têtes correspondent aux titres et aux sous-titres des sections d'un contenu disponible sur une page web. Ils renseignent les lecteurs, y compris les robots, sur la structure du contenu.

- Utiliser une seule balise <H1> par page. Elle s'applique généralement au titre dans le contenu même de la page, mais il n'est pas nécessaire qu'elle soit identique à la balise de titre.
- Ajouter les mots-clés et des termes liés aux en-têtes, en nombre raisonnable, pour aider Google à interpréter le sujet traité dans le contenu de la page.

Problèmes liés aux URL

L'essentiel des bonnes pratiques en matière de création d'URL ont été pensées pour aider les visiteurs humains à mieux les comprendre. Elles recommandent notamment de privilégier la concision, d'utiliser des termes compréhensibles séparés par des tirets ou encore d'ajouter des mots-clés.

Si ces recommandations semblent davantage s'appliquer à l'optimisation de l'expérience utilisateur, elles n'en demeurent pas moins importantes pour le SEO. Bien qu'elles ne permettent pas d'améliorer les chances d'indexation d'une URL, elles sont susceptibles de booster son taux de clics.

Google formule même quelques conseils dédiés exclusivement au SEO technique :

- Utiliser une structure d'URL simple.
- Éviter d'utiliser un trop grand nombre de paramètres d'URL. Ils peuvent générer une quantité d'URL conséquente et desservent les robots d'exploration. Ce problème est également fréquent dans le cadre de la navigation à facettes, qui permet aux utilisateurs d'appliquer un ou plusieurs filtres à une page pour la personnaliser.
- Éviter d'utiliser le symbole dièse dans une URL. À l'origine, il servait de point d'ancrage pour des sections du contenu et Google considère toujours les URL affichant les mêmes caractères avant ce symbole comme une seule et même page.
- Une technologie comme AJAX autorise l'utilisation du hashbang (#!) au lieu du symbole dièse.

De manière générale, il est conseillé de ne pas modifier une URL, sauf en cas d'absolue nécessité.

Un audit technique doit permettre de répondre aux questions suivantes :

- Les URL contiennent-elles trop de paramètres ?
- Les robots de Google consacrent-ils trop de temps à l'exploration des pages en mode navigation à facettes ?
- Les URL à indexer contiennent-elles le symbole dièse (#) ?
- La structure des URL est-elle simple et facile à lire ?

La taille des contenus multimédias

Le contenu multimédia d'une page influence ses performances et plus particulièrement son temps de chargement. Ils comprennent les images, vidéos, fichiers .css, JavaScript, .pdf et de polices, ainsi que tous les autres fichiers devant être téléchargés pour permettre l'affichage de la page. Lorsque ces fichiers sont trop volumineux ou trop nombreux, ils peuvent ralentir le temps de chargement.

Il existe différentes techniques pour réduire la taille d'une page :

- Supprimer les ressources téléchargées qui ne sont pas utilisées
- Compresser les données
- Remplacer les polices intégrées par des polices web
- Optimiser les images en réduisant leur taille, par exemple en remplaçant les images par des effets .css, en utilisant les résolutions et formats d'image appropriés, en compressant les images ou encore en utilisant le chargement différé.
- Remplacer les .gif par des vidéos

Un audit technique doit :

- Utiliser les données sur la taille, la vitesse et les performances d'une page pour déterminer le poids à partir duquel elle devient trop volumineuse pour obtenir de bonnes performances. OnCrawl définit par exemple les poids et temps de chargement idéaux d'une page comme étant respectivement inférieurs à 100 Ko et 500 ms.
- Identifier les images, vidéos et autres contenus multimédias particulièrement volumineux.

La longueur du contenu

Définir la longueur idéale d'un contenu pour le SEO est un débat continu. Certains estiment qu'elle ne devrait pas dépasser quelques centaines de mots, alors que pour d'autres, le contenu supérieur à 2 000 mots offre de meilleures performances. La longueur pourrait même n'avoir aucune importance.

L'audit technique d'un site doit vérifier s'il existe une corrélation entre le nombre de mots d'un contenu on-page, quel que soit son type, et :

- La profondeur de la page
- Les pages classées dans les résultats de recherche

Chaque site web et chaque contenu est différent. Pour définir la longueur idéale d'un contenu on-page, il est donc nécessaire de le comparer aux mêmes types de contenus en sélectionnant ceux qui obtiennent les meilleures performances.

Quelle que soit sa longueur, un contenu doit être bien rédigé et répondre aux objectifs qui lui ont été fixés.

La rédaction de contenu on-page performant s'articule autour de trois axes :

- Créer des contenus courts et longs, car ils peuvent répondre à des besoins différents.
- Appliquer les principes de rédaction suivants :
 - Privilégier la concision et la simplicité.
 - Comparer le contenu du site à celui d'autres sites ou blogs de qualité. Tirer de nouvelles idées à tester de cette analyse comparative et évaluer leurs performances.
- Mesurer l'impact de la longueur d'un contenu sur ses performances dans le cadre d'un audit et l'ajuster en fonction des résultats obtenus.

Enfin, il est important de noter que la qualité rédactionnelle d'un contenu et le choix de l'emplacement des balises HTML, qui permet de l'optimiser pour les robots d'exploration, jouent un rôle central dans ses performances.

Sources :

Méta-descriptions : https://support.google.com/webmasters/answer/35624?visit_id=636758086798903067-3237243957&rd=1%20&hl=fr

Balises méta : <https://fr.oncrawl.com/referencement/balises-metas-impact-seo/>

Informations spécifiques à la balise de titre : <https://fr.oncrawl.com/referencement/12-choses-a-savoir-balises-titre/>

Longueur de contenu : <https://blog.hubspot.com/marketing/how-long-should-your-blog-posts-be-faq>

Le contenu dupliqué



Le contenu dupliqué

Lorsque plusieurs URL présentent du contenu identique, ce contenu est dit dupliqué.

Cela peut se produire lorsque :

- Le contenu d'une page est insuffisant par rapport aux autres éléments (l'en-tête, le pied de page, le menu latéral et toute autre information relative à la structure de la page), même si de nombreux moteurs de recherche leur accordent de moins en moins d'importance et préfèrent se concentrer sur le contenu en lui-même.
- Des textes sont réutilisés comme en-têtes, titres de pages ou méta-descriptions. Le même extrait descriptif peut alors être publié dans les résultats de recherche pour deux pages qui sont en réalité différentes.
- Deux versions identiques d'une même page sont disponibles sur un même site, qu'il s'agisse d'un doublon ou d'une page de test.
- Le site présente deux pages similaires, par exemple la page produit d'un clou de 15 mm et celle d'un clou de 17 mm, qui ne se distinguent que par des différences minimales, comme ici la longueur en millimètres dans le nom de chaque produit.

Il peut également arriver que du contenu dupliqué soit créé lorsque le CMS, le système de gestion de contenu d'un site (WordPress, par exemple), utilise différentes formules pour faire référence à une même page.

Il est important de noter que si les quatre URL ci-dessous permettent bien d'accéder à une seule et même page, les moteurs de recherche considèrent qu'il en existe quatre distinctes redirigeant vers le même contenu :

- <https://www.exemple.com/a-propos/>
- <http://www.exemple.com/a-propos/>
- https://www.exemple.com/a-propos?utm=mес_parametres
- <https://www.exemple.com/categorie/article/798/>

Les problèmes liés au contenu dupliqué

Le problème fondamental posé par le contenu dupliqué est simple : comment les robots d'exploration identifient-ils les URL à indexer ?

Lorsque différentes URL redirigeant vers une même page sont indexées, elles se font concurrence tant au niveau du classement dans les résultats de recherche que des taux de clics, alors que l'utilisation d'une URL canonique favoriserait un meilleur référencement.

La recherche de contenu dupliqué

Dans le cadre d'un audit technique, une entreprise devrait utiliser un outil tel qu'un robot pour identifier le contenu dupliqué sur son site.

Les types de contenu dupliqué ou quasi dupliqué à prendre en compte sont les suivants :

- Les titres de pages
- Les balises H1
- Les balises H2
- Les méta-descriptions
- Les textes complets des pages

Les solutions face au contenu dupliqué : des pages distinctes proposant le même contenu

Dans le cas d'éléments redondants dans des URL redirigeant vers des pages distinctes, la meilleure solution consiste à modifier ces éléments pour les rendre uniques.

- Les titres de pages
- Les balises H1
- Les balises H2
- Les méta-descriptions
- Les textes complets de pages distinctes associées à différentes URL, comme les pages produit des clous prises en exemple en début de chapitre

Un audit technique doit permettre de dresser une liste des pages affectées par la redondance de contenu, des types de problèmes rencontrés et des pages où sont publiés les mêmes contenus.

Il est important de vérifier les modifications effectuées sur une page affichant du contenu dupliqué pour s'assurer que le problème est résolu.

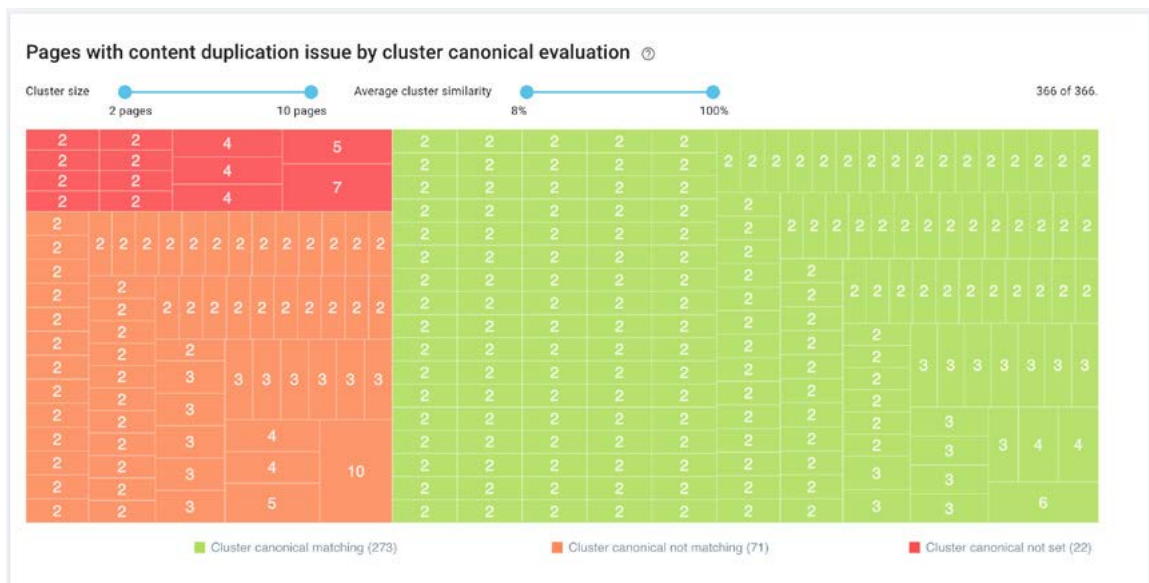
Les solutions face au contenu dupliqué : plusieurs URL redirigeant vers une même page

Lorsque du contenu dupliqué est généré par plusieurs URL redirigeant vers une même page, la meilleure stratégie consiste à préciser aux moteurs de recherche laquelle doit être considérée comme l'URL canonique.

Pour ce faire, il suffit d'ajouter l'élément `<link rel="canonical">` dans l'en-tête de l'URL sélectionnée. Cette opération permet d'indiquer qu'il s'agit de l'URL devant être indexée parmi toutes celles qui redirigent vers la même page.

Un audit technique doit inclure une liste d'URL canoniques, ainsi que les URL vers lesquelles elles redirigent. À cette fin, il doit permettre de :

- Vérifier qu'une URL canonique a bien été définie pour toutes les pages accessibles depuis plusieurs URL.
- Vérifier que toutes les URL redirigeant vers une même page possèdent la même URL canonique.
- Vérifier qu'il existe une version mobile d'une URL canonique en ajoutant l'élément `<link rel="alternate">` dans l'en-tête de cette dernière.



Utilisation d'OnCrawl pour la gestion de groupes de contenus redondants

Sources :

<https://webmasters.googleblog.com/2013/04/5-common-mistakes-with-relcanonical.html>

<https://support.google.com/webmasters/answer/139066?hl=fr>

<https://moz.com/blog/rel-canonical>

Les problématiques liées aux appareils mobiles

10

Les problématiques liées aux appareils mobiles

Les recherches sur mobile représentent une part de plus en plus importante de l'ensemble des recherches effectuées en ligne. Dès 2015, Google indiquait d'ailleurs qu'elles dépassaient le nombre de recherches sur ordinateur. À l'été 2018, ce nombre atteignait même 60 %, incitant la plateforme à commencer l'indexation de pages mobiles en septembre de la même année.

Non seulement le nombre de visites et de recherches depuis les appareils mobiles augmente, mais Google utilise aussi des robots d'exploration qui leur sont exclusivement destinés. Les pages d'un site qui s'affichent rapidement et correctement sur mobile sont donc susceptibles d'être plus performantes.

Aujourd'hui, la recherche mobile doit tenir compte de :

- L'obtention d'un classement élevé dans les résultats de recherche, grâce à l'optimisation de l'espace disponible sur les écrans des appareils mobiles et le recours de plus en plus fréquent à la recherche vocale.
- L'indexation des pages mobiles des sites proposant des expériences différentes sur ordinateur et sur mobile.
- L'importance de la vitesse de chargement des pages.
- La capacité à proposer des pages AMP.
- Des classements dans les résultats de recherche supérieurs de 80 % à ceux enregistrés sur ordinateur.

L'un des enjeux d'un audit technique consiste à déterminer si un site est correctement configuré pour une utilisation mobile fluide à l'aide d'un [outil de test d'optimisation mobile](#), comme celui proposé par Google.

L'indexation orientée mobile

En 2018, Google a commencé à indexer les pages mobiles plutôt que leur version pour ordinateur. L'extrait descriptif d'une page indexée, y compris son titre, sa description et la version mise en cache, peut donc désormais provenir de la version mobile d'un site. Celle-ci devient alors le principal contenu utilisé dans le classement d'une URL.

Pour les sites qui utilisent des pages au format AMP, l'indexation orientée mobile privilégiera les versions mobiles standards plutôt que les versions AMP.

Les conseils suivants prodigués par Google sont particulièrement utiles lorsqu'un site se compose de pages distinctes et dynamiques sur mobile :

- Utiliser des contenus identiques sur les versions mobile et bureau du site.
- S'assurer que les données structurées sont présentes dans les deux versions du site.
- S'assurer que les éléments méta sont intégrés aux deux versions du site.

Le choix de l'indexation orientée mobile s'accompagne d'un pic d'exploration, car le robot de Google dédié aux appareils mobiles parcourt toutes les URL connues du site pendant plusieurs jours, voire plusieurs semaines. En contrepartie, l'exploration de la version bureau du site tend à diminuer. Ces tendances se manifestent sous la forme d'un pic dans la Search Console de Google. Elles sont également perceptibles en assurant le suivi des visites du robot de Google dans les fichiers journaux du serveur.

Un audit technique doit :

- Vérifier le caractère explorable et indexable des pages mobiles.
- Utiliser les données des fichiers journaux du serveur pour savoir si l'indexation orientée mobile est effective.
- S'assurer que les éléments `<link rel="alternate">` et `<link rel="canonical">` sont correctement implémentés sur le site pour éviter tout problème de redondance de contenu sur l'une ou l'autre de ses versions.
- S'assurer que tous les éléments `<link rel="hreflang">` redirigent vers les bonnes pages : les éléments `<link rel="hreflang">` des pages mobiles doivent rediriger vers des pages mobiles et ceux des pages bureau vers le même type de page.
- Vérifier que les instructions du fichier robots.txt sont identiques pour les versions mobile et bureau du site, sauf lorsque des exceptions sont nécessaires.

Le design adaptatif

Le design adaptatif permet à une page web d'adapter l'affichage de son contenu à la taille de l'écran d'un appareil mobile. Outre l'impact positif sur l'expérience utilisateur, cette technique permet également de n'utiliser qu'une seule et même page sur les versions mobile et bureau du site.

Elle est souvent considérée comme l'une des bonnes pratiques en matière de SEO, car elle entraîne une diminution de la charge de travail autrement nécessaire à l'optimisation et à l'indexation des pages mobiles d'un site.

Google AMP : Accelerated Mobile Pages

Il s'agit d'une technologie open source soutenue par Google et Twitter, accessible à tout type d'applications, de web viewers et de navigateurs. Google, Twitter, LinkedIn et Reddit font partie des plateformes qui l'utilisent actuellement. Son principal objectif consiste à créer un format de page offrant une expérience mobile plus rapide.

Elle se compose de HTML, de fichiers JavaScript et d'un système de cache qui, grâce à des extensions et à des propriétés propres au format AMP, accélèrent le temps de chargement des pages mobiles, même lorsqu'elles comprennent des contenus enrichis comme des vidéos, des fichiers .pdf ou audio, ou encore des infographies.

Le format AMP repose sur un codage léger et l'hébergement de contenu dans Google AMP Cache. Ce cache permet de précharger les pages AMP et de les afficher quasi instantanément dans les résultats de recherche, tout en garantissant la conformité de leur format. Elles sont facilement identifiables grâce à l'éclair bleu situé à côté de leur URL.

Les pages AMP présentent les avantages suivants :

- Elles s'affichent correctement sur tout type d'appareils.
- Elles bénéficient d'un hébergement fiable sur Google.
- Elles offrent un chargement rapide adapté aux connexions de données mobiles et aux processeurs d'appareils mobiles.
- Elles sont en passe de devenir la norme dans les pages de résultats de Google.

Utiliser une technique comme le web scraping dans le cadre de l'exploration d'un site permet d'identifier les pages qui existent au format AMP.

Un audit technique doit :

- Vérifier que le site comprend des pages AMP.
- Vérifier que ces pages s'affichent correctement.

Sources :

<https://adwords.googleblog.com/2015/05/building-for-next-moment.html>

L'indexation orientée mobile : <https://webmasters.googleblog.com/2018/03/rolling-out-mobile-first-indexing.html>

Les bonnes pratiques de l'indexation orientée mobile : <https://developers.google.com/search/mobile-sites/mobile-first-indexing?hl=fr>

AMP : <https://www.oncrawl.com/oncrawl-seo-thoughts/google-amp-project-what-to-know/>

AMP : <https://fr.oncrawl.com/referencement/passage-amp-bonnes-pratiques/>

AMP : <https://developers.google.com/search/docs/guides/about-amp?hl=fr>

Les problématiques liées aux sites internationaux



Les problématiques liées aux sites internationaux

Un site international fournit des contenus adaptés à différentes régions ou en différentes langues.

Cet e-book se concentre sur ce dernier aspect, bien que tout contenu proposé dans une seule langue puisse tout de même présenter des nuances dues à des différences de coutumes, de réglementations ou de devises.

La structure d'un site international

Le contenu d'un site international peut s'organiser de différentes façons. Google recommande de ne pas recourir à l'analyse d'adresse IP pour rediriger les visiteurs vers la version localisée du site qui leur correspond.

Quelle que soit la structure d'un site, un audit technique doit :

- S'assurer qu'elle permet à Google d'explorer et d'indexer toutes les pages. Pour les sites qui offrent du contenu dynamique ou qui redirigent automatiquement les visiteurs vers une page localisée depuis la page d'accueil, il est important de noter que les paramètres de langue de la plupart des robots de Google sont définis sur l'anglais américain.
- Analyser la structure des liens internes pour vérifier qu'ils ne redirigent pas vers des pages s'affichant automatiquement dans une autre langue.
- Vérifier que la langue utilisée sur une page est facile à identifier. Google s'attend à ce qu'il n'y en ait qu'une par page. Les pages dont le menu de navigation et le contenu sont proposés dans deux langues différentes, ou celles qui affichent à la fois un contenu en version originale et sa traduction, peuvent donc être source de confusion pour les robots de la plateforme.

L'organisation pêle-mêle

L'organisation pêle-mêle fait référence aux sites où aucun élément ne permet de dissocier les paramètres de langue des pages. Ces sites ne sont pas optimisés pour la recherche et entraînent généralement quelques difficultés navigationnelles pour Google et les visiteurs.

Un audit technique doit donc permettre de classer toutes les pages d'un site par langue.

Les paramètres d'URL

Chaque langue proposée sur un site est identifiable grâce à un paramètre d'URL tel que **https://www.exemple.com?loc=fr**.

L'utilisation de ce paramètre est toutefois déconseillée car il peut poser des problèmes aux robots d'exploration, aux outils de monitoring de certains sites et aux visiteurs.

L'organisation en répertoires

Chaque langue est stockée dans son propre répertoire, par exemple **https://www.exemple.com/fr** ou **https://www.exemple.com/en**. Ce type d'organisation constitue l'une des meilleures stratégies pour un site disponible en plusieurs langues.

Un audit technique doit donc permettre de s'assurer que tous les répertoires sont accessibles aux robots de Google.

L'organisation en sous-domaines

Chaque langue est stockée dans son propre sous-domaine, par exemple **https://fr.exemple.com** ou **https://en.exemple.com**.

Un audit technique doit :

- Inclure les pages de tous les sous-domaines de langues.
- S'assurer que la configuration des sous-domaines permet aux robots de Google d'y accéder.

L'organisation en domaines

Chaque langue possède son propre domaine, par exemple **https://www.exemple.co.uk** ou **https://www.exemple.fr**. Ce type d'organisation est conseillé aux sites ciblant différents pays.

Ici, un audit technique doit tenir compte de tous les sites pour chaque langue.

La gestion de contenu disponible en plusieurs langues

Google cherche à définir la langue utilisée sur une page en fonction de différents facteurs :

- Les informations disponibles dans le Rapport sur le ciblage international de la Search Console.
- Le niveau national du domaine (<https://www.exemple.fr>), le cas échéant.
- La valeur de l'attribut hreflang.
- La localisation du serveur.
- Les informations on-page, telles que l'adresse, le numéro de téléphone ou la devise.

Il est essentiel que l'audit technique d'un site international tienne compte de la valeur de l'attribut hreflang et de son implémentation.

Cette valeur est importante pour plusieurs raisons :

- Elle permet à Google d'identifier correctement la langue d'un contenu.
- Elle permet de limiter la création de contenu dupliqué en cas de contenu disponible dans la même langue pour deux pays différents.
- Elle permet de conférer la notoriété d'un contenu existant à sa traduction.

La valeur de l'attribut hreflang, qui identifie les versions alternatives d'une URL (ses traductions), se présente sous la forme d'un code de langue à deux lettres au format ISO 639-1, et d'un code de pays facultatif au format ISO 3166-1 Alpha 2. Elle génère une liste des liens redirigeant vers la version originale d'une page et vers toutes ses traductions, où qu'elles se trouvent sur internet.

Pour éviter que des traductions non autorisées publiées sur un autre site ne soient référencées comme les URL alternatives d'une page, la valeur de l'attribut hreflang doit être bidirectionnelle : si la page A définit la page B comme traduction, la page B doit en faire autant pour la page A.

La valeur hreflang="default-x" peut être utilisée pour les pages génériques non traduites ou pour les pages de destination qui ont pour seule vocation la sélection d'une langue ou d'une région.

La valeur de l'attribut hreflang peut être implémentée :

- Dans la section **<head>** d'une page, avant tout élément susceptible de fermer prématurément la section. Pour une traduction en anglais et en français, la valeur de l'attribut se présente alors ainsi :

```
<link rel="alternate" hreflang="en" href="https://www.exemple.com/en/">
<link rel="alternate" hreflang="fr" href="https://www.exemple.com/fr/">
```

- Dans les **en-têtes HTTP** pour les contenus qui n'utilisent pas de code HTML, comme ici :

```
Link: <https://www.exemple.com/en/>; rel="alternate"; hreflang="en",
<https://www.exemple.com/fr/>; rel="alternate"; hreflang="fr"
```

- Dans un **fichier sitemap.xml**. La valeur de l'attribut doit être définie pour toutes les URL qui correspondent aux traductions d'un même contenu, y compris pour celui-ci. Toutes ces URL doivent également être enregistrées dans le même fichier sitemap. La valeur de l'attribut se présente alors ainsi :

```
<url>
  <loc>http://www.exemple.com/en/</loc>
  <xhtml:link rel="alternate" hreflang="fr" href="http://www.exemple.com/
fr/" />
  <xhtml:link rel="alternate" hreflang="en" href="http://www.exemple.com/
en/" />
</url>
```

Un audit technique doit :

- S'assurer que la valeur de l'attribut hreflang a été définie.
- S'assurer qu'elle est intégrée à une URL complète (<https://www.exemple.com/fr/>).
- S'assurer que la valeur de l'attribut est définie pour toutes les URL qui correspondent aux traductions d'un même contenu, y compris pour le contenu lui-même.
- S'assurer que toutes les valeurs redirigent vers des pages qui existent.

Sources :

<https://support.google.com/webmasters/answer/182192?hl=fr>

<https://support.google.com/webmasters/answer/189077?hl=fr>

<https://fr.oncrawl.com/referencement/hreflang-seo-5-erreurs-eviter/>

<https://fr.oncrawl.com/product-update/metriques-hreflang-groupes-pages-traduites/>

Schema.org

12

Schema.org

Les données structurées sont des balises qui aident les moteurs de recherche à traiter les contenus en tant qu'entités, afin de définir plus facilement les relations qui les unissent les uns aux autres. Elles permettent aux robots d'exploration et à d'autres algorithmes de reconnaître certains types de contenus et de proposer des résultats plus pertinents.

[Schema.org](https://schema.org) définit les normes relatives aux données structurées.

Google utilise les données structurées pour générer :

- Des résultats sous forme de fiches info
- Des carrousels de vidéos, d'images, d'ingrédients, d'événements populaires
- Des résultats relatifs à l'actualité
- Des résultats cartographiques
- Des extraits descriptifs plus informatifs dans les pages de résultats. Ceux-ci peuvent comprendre un classement par étoiles, les horaires de diffusion d'un film, des tarifs, le nombre de calories d'une recette, une zone de recherche, des coordonnées professionnelles, etc.

Ce type d'extrait descriptif permet à une URL de se distinguer des autres sur une page de résultats, d'accéder à des positions convoitées comme celle située sur la droite des résultats naturels, et même d'être sélectionnée comme extrait optimisé.

Les avantages de ces extraits enrichis dépendent du type d'informations fournies et de la façon dont elles sont présentées dans les pages de résultats des moteurs de recherche. Dans certains cas, ils permettent d'améliorer les taux de clics ou de cibler une audience plus large.

Voici une liste de données structurées :

Prises en charge par Google :

- Fil d'Ariane
- Coordonnées professionnelles
- Carrousel
- Logo
- Zone de recherche
- Profil sur les réseaux sociaux
- Article
- Livre
- Formation
- Ensemble de données
- Événement
- Fact check (vérification d'informations)
- Offre d'emploi
- Entreprise locale
- Contenu multimédia (accès limité)
- Profession
- Podcast
- Produit
- Recette
- Avis
- Application logicielle (version beta)
- Recherche vocale (version beta)
- Liste de lieux populaires (version beta)
- Vidéo

Prises en charge par Bing :

- Fil d'Ariane
- Entreprises et organisations
- Événements
- Personnes
- Produits et offres
- Recettes
- Avis
- Avis (regroupés)
- ClaimReview / Fact check

Un audit technique doit :


- Identifier les pages avec données structurées.
- Identifier le type de données structurées utilisées sur le site.
- Vérifier la validité des données structurées utilisées.
- Comparer les performances (classement dans les pages de résultats des moteurs de recherche, visites de robots, visites naturelles) des pages avec données structurées à celles des autres pages, afin de définir la valeur de ce type de données pour le site.

Voici quelques exemples :

Google restaurant lyon




Tous Maps Actualités Images Shopping Plus Paramètres Outils

Environ 145 000 000 résultats (0,59 secondes)



Note - Cuisine - Horaires -

Toussaint : Ces horaires sont susceptibles de varier.

<p>Le Ti'Punch de la Réunion</p> <p>4,4 ★★★★★ (107) - €€ - Restaurant</p> <p>12 Rue Longue</p> <p>Fermé · Ouvre à 19:30</p> <p>Spécialités réunionnaises</p>	
<p>Restaurant Le Musée</p> <p>4,7 ★★★★★ (792) - €€ - Française</p> <p>2 Rue des Forces</p> <p>Fermé · Ouvre à 12:00</p> <p>Lieu chaleureux · Décontracté · Groupes</p>	
<p>RESTAURANT Takao Takano</p> <p>4,7 ★★★★★ (346) - €€€ - Restaurant</p> <p>33 Rue Malesherbes</p> <p>Plats sophistiqués & déco contemporaine</p>	

☰ Autres adresses



Chat

Animal

Le Chat domestique est la sous-espèce issue de la domestication du Chat sauvage, mammifère carnivore de la famille des Félidés. Il est l'un des principaux animaux de compagnie et compte aujourd'hui une cinquantaine de races différentes reconnues par les instances de certification. [Wikipédia](#)

Famille : [Felidae](#)

Espérance de vie : 2 – 16 ans (À l'état sauvage)

Ordre : Carnivora

Période de gestation : 58 – 67 jours

Règne : Animalia

Poids : 3,6 – 4,5 kg (Adulte)

Races

[Voir d'autres éléments \(plus de 15\)](#)



Persan



Bengal



Maine
coon



Siamois



British
shorthair



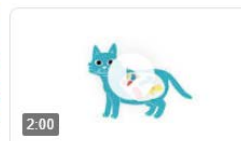
Un chiot profite du handicap d'un chat

Koreus - Il y a 2 jours



Vous souhaitez gronder votre chat ? Comment lui faire ...

LaVieDesChats
YouTube - 24 mars 2019



Born to survive, le chat - Le proliférateur

Arte - 28 févr. 2019

Sources:

Documentation (Google) : <https://support.google.com/webmasters/answer/7358659?hl=fr>

Documentation (Bing) : <https://www.bing.com/webmaster/help/marking-up-your-site-withstructured-data-3a93e731>

Tirer les conclusions d'un audit

13

Tirer les conclusions d'un audit

Les étapes qui suivent la réalisation d'un audit technique sont cruciales pour mettre à profit les efforts qu'il a engagés.

Un audit n'a pas de grande valeur s'il n'aboutit pas à des solutions concrètes et à l'implémentation des mises à jour nécessaires. Ce processus s'articule autour de trois axes :

Définir les actions à mettre en place

Il s'agit en premier lieu d'identifier les erreurs et les leviers d'amélioration décelés par un audit afin d'y apporter des solutions et de les implémenter. Si l'audit d'un site révèle qu'il génère des messages d'erreur 404 de façon récurrente parce qu'un trop grand nombre de liens internes sont obsolètes, il faudra par exemple rediriger les pages affectées vers des pages opérationnelles, afin qu'elles soient mieux référencées dans les résultats de recherche. Le même principe s'applique à toutes les erreurs que l'audit permet de découvrir.

Définir un ordre de priorité

Une fois les actions à mettre en place définies, il faut les prioriser en fonction des ressources et des objectifs de l'entreprise. Par exemple, la mise à jour d'un lien cassé sur la page d'accueil représente une solution à fort impact qui demande peu de ressources. En revanche, si la restructuration d'un site peut, elle aussi, avoir un impact positif sur ses performances et son classement dans les résultats de recherche, elle nécessitera davantage de ressources. Le processus de priorisation permet de cibler les actions les plus rentables pour l'entreprise.

Être attentif aux objectifs commerciaux

Enfin, il est essentiel que les actions à mettre en place s'alignent sur les objectifs de l'entreprise. Par exemple, définir une stratégie de groupe thématique pour réorganiser le blog n'est pas la seule solution pour qu'il génère davantage de trafic. Supprimer les contenus redondants ou optimiser la stratégie de recherche de mots-clés sont d'autres solutions permettant d'aboutir au même résultat. Les objectifs commerciaux d'une entreprise éclairent le choix des actions à mettre en place, en ciblant celles qui permettront d'obtenir les résultats les plus significatifs.

Conclusion

14

Conclusion

La mise en pratique

Cet e-book constitue un guide permettant de réaliser un audit technique, d'identifier les problèmes et les améliorations possibles qu'il vise à mettre en évidence, et de définir les actions à implémenter pour qu'un site génère davantage de trafic.

La réalisation d'un audit technique ne doit pas être une initiative ponctuelle. Il s'agit d'un processus récurrent qui permet de suivre les performances d'un site à long terme, afin de parfaire sa stratégie d'optimisation. Il est généralement conseillé d'auditer un site au moins une fois par an, ou suite à toute mise à jour de sa structure ou de son design.

The HubSpot logo is displayed in white on an orange background. It features the word "HubSpot" in a sans-serif font, with a stylized robot head icon integrated into the letter "o".

Faites vos premiers pas avec HubSpot

Grâce à des outils qui permettent d'organiser à grande échelle et de façon personnalisée chaque étape de votre processus et à une équipe de support prête à vous aider, mettre en place la méthodologie inbound n'a jamais été aussi facile.

[Créer un compte gratuit](#)

The OnCrawl logo is displayed in white on a blue background. It features the word "OnCrawl" in a sans-serif font, with a stylized robot head icon integrated into the letter "o".

Unleash your SEO potential

La plateforme de SEO technique qui aide les entreprises à prendre des décisions plus éclairées en réalisant des audits SEO et en assurant le suivi de leur site.

[Essayer gratuitement](#)

HubSpot



OnCrawl

