

Explication des statistiques Webalizer

Sources :

<http://www.eolas.fr/8-conseil/65-interpreter-vos-statistiques-webalizer.htm>

<http://support.sherweb.com/faqDetails.php?IDArticle=68>

Un site web est un ensemble de documents HTML et de fichiers, scripts et bases de données associés, qu'un serveur HTTP met à disposition des internautes sur le World Wide Web.

Pour accéder à un site web, l'utilisateur a besoin d'un navigateur et d'une connexion à l'Internet.

Lorsqu'un internaute consulte un site à partir d'un navigateur, son ordinateur lance des requêtes HTTP au serveur et reçoit ses réponses, généralement sous forme de pages. Il est alors possible d'analyser l'origine de ces consultations.

A quoi sert l'analyse des statistiques de consultation de votre site ?

L'objectif de l'analyse des statistiques est de connaître et d'analyser la façon dont les internautes ont accédé à votre site et l'ont visité. Vous pouvez également savoir quelle consommation de trafic (Mb transférés) génère l'utilisation de votre site.

Le suivi d'indicateurs clés (pages les plus consultées, erreurs d'affichage, ...) permet d'orienter vos choix d'optimisation de votre site en termes d'ergonomie et de contenu.

Quelles données exploiter ?

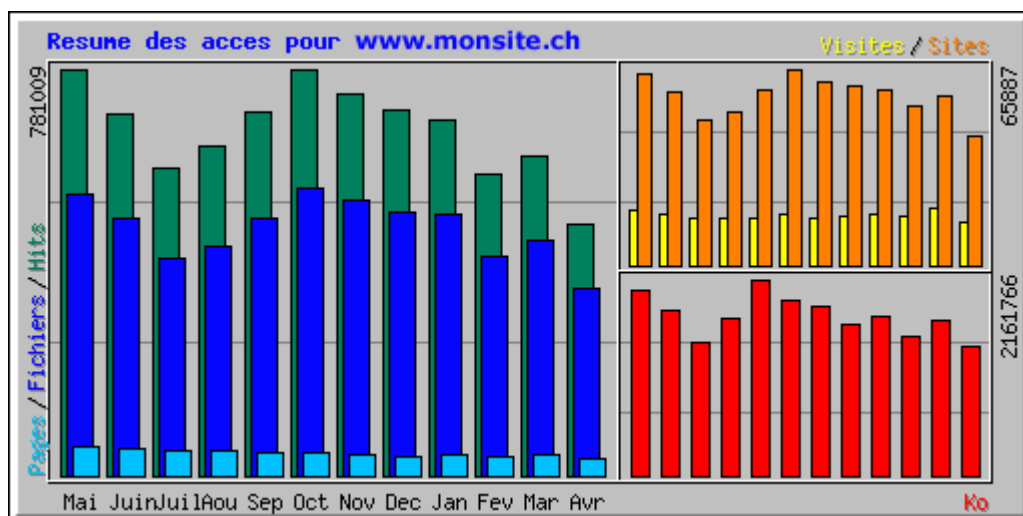
De manière générale, deux types de données statistiques peuvent être exploités, celles :

- sur la navigation des internautes (fréquentation, pages visitées,...) utilisées pour optimiser l'ergonomie et le contenu du site,
- plus techniques (transfert mensuel, serveurs,...) plutôt utilisées au niveau du développement du site.

Description de vos statistiques Webalizer

1° Évolution annuelle, mensuelle et journalière

Vous avez accès à un état annuel de vos statistiques sur la première page sous forme de tableaux et de graphiques.



Résumé par mois										
Mois	Moyenne journalière				Totaux mensuels					
	Hits	Fichiers	Pages	Visites	Sites	Ko	Visites	Pages	Fichiers	Hits
Avr 2005	15966	14063	2101	985	21814	3947312	26612	56739	379708	431095
Mar 2005	14851	13031	1980	902	23651	4221468	27981	61395	403961	460382
Fev 2005	14488	12752	1904	860	21044	3710988	24104	53335	357072	405689
Jan 2005	12021	10659	1629	749	19324	3371171	23239	50508	330444	372681
Dec 2004	10028	8927	1437	635	16261	2935603	19688	44559	276756	310874
Nov 2004	10904	9693	1452	641	16541	3160422	19238	43563	290810	327128
Oct 2004	9215	8181	1292	565	15651	2859926	17545	40066	253623	285670
Sep 2004	7727	6784	1143	496	12608	2201968	14892	34310	203528	231819
Aou 2004	6258	5451	1015	421	11323	1862737	13055	31487	168999	194008
Juil 2004	6009	5204	914	367	10428	1704068	11390	28349	161349	186301
Juin 2004	8175	7068	1205	417	10173	2135218	12530	36155	212045	245277
Mai 2004	6464	5567	1002	402	10101	1836240	12491	31066	172603	200413
Totaux						33947121	222765	511532	3210898	3651337

- Le graphique porte sur une période de 1 an et vous permet de visualiser l'évolution des connexions à votre site durant l'année en cours.
- Une analyse statistique se fait d'un mois à l'autre afin de déterminer l'évolution de la consultation de votre site Internet.
- Les liens sur chaque mois vous permettent d'accéder aux détails des statistiques mensuelles, journalières et même horaires.
- Les statistiques de connexion au site sont mises à jour tous les soirs.

2° Vos statistiques

Les statistiques sont réparties en plusieurs rubriques. Chacune d'elles fournit des indications utilisant certains termes techniques. Ces termes sont définis au paragraphe "Définition des principaux termes utilisés" ci-dessous.

Statistiques mensuelles

Détaille les totaux des pages, hits, etc.

- Nombre de pages vues : permet de connaître la densité de pages visitées
- Nombre de visites : le nombre de sessions différentes (un même visiteur peut venir plusieurs fois dans le mois)

Hits (ou requêtes) par codes de réponses

Totaux et moyennes pour tous les codes de réponse du serveur. Cela permet de connaître le nombre de fichiers transférés (200), le nombre de requêtes sur un fichier qui n'existe pas (404), le nombre de téléchargements interrompus (206), etc. (c.f. la liste détaillée des codes en fin de document)

Statistiques journalières et statistiques horaires

Totaux et moyennes des valeurs jour par jour et heure par heure.

URLs

La rubrique URL vous permet de visualiser les pages de votre site les plus souvent consultées.

Chacun des fichiers du site web correspond à un URL. Il existe deux tableaux. Le premier est une classification par nombre de requêtes, le second est une classification en fonction de la quantité d'octets transférés.

Pages d'entrées et Pages de sortie

Ces rubriques vous permettent de savoir comment les visiteurs utilisent votre site et y circulent : «Pages d'entrées» donne la référence de la page sur laquelle les utilisateurs se sont connectés et «Pages de sortie» celle à partir de laquelle ils ont quitté le site.

Sites

Permet de repérer les visiteurs de votre site. Il s'agit des machines clientes ; elles sont identifiées par leurs noms ou leurs numéros IP. Comme pour les URL, il y a deux classifications. La première en fonction du nombre de requêtes, l'autre en fonction du nombre d'octets transférés.

Ko

Nombre de KBytes transféré sur le serveur, permet d'optimiser le poids des pages.

Référents

Fournit une estimation approximative de quels URL's sont utilisées pour entrer dans le site ou pour accéder à une page du site (c'est pourquoi on trouve ici des pages du site qui ont permis d'accéder à d'autres pages du même site). À cause des limitations dans le protocole HTTP, des rotations de log, etc... ce chiffre doit être considéré comme une «approximation brute» des nombres réels, nous obtenons cependant une bonne indication de la tendance globale.

Mots-clés

Mots-clés utilisés dans les moteurs de recherche qui ont servi à trouver votre site

Navigateurs

Différents navigateurs utilisés par les internautes qui visitent votre site

Pays

Domaines qui ont visité le site (.fr .com .net .us etc ...)

3° Définition des principaux termes utilisés

Hits

Un Hit est une requête d'une machine cliente sur le serveur (si dans une page, il y a des images incluses, il y aura plusieurs requêtes effectuées pour afficher la page : 1 pour le fichier html, 1 par image, etc.). Ces requêtes peuvent être des pages html, des fichiers graphiques, des fichiers audio, etc...

Ainsi, une visite correspond à plusieurs hits si la page visualisée contient de nombreux éléments.

° Remarque : La relation entre un "hit" et un "fichier" pourrait être interprétée comme celle entre une "requête entrante" et une "réponse sortante".

Files ou fichiers

Cette valeur comptabilise les transferts effectifs de fichiers du serveur vers le client. Le Fichier peut être de n'importe quel type (html, images, pdf, doc, ..). Nous sommes ici dans le cas où les données ne sont pas dans un cache et doivent être effectivement transmises. Pour les techniciens, cela correspond à toutes les requêtes HTTP identifiées avec un code 200.

Concrètement, c'est la récupération, depuis le site web, d'un document tel qu'une page d'accueil ou une image. Le nombre de fichiers est toujours inférieur ou égal au nombre de hits. Par exemple, une page qui contient deux images et un texte renvoie 4 hits : un pour la page et un pour chaque objet contenu dans la page. Par contre, pour la même page, le nombre de fichiers est de 3, les blocs de texte n'étant pas considérés comme des fichiers.

KBytes et KOctets

Ces valeurs montrent la quantité de données qui a été envoyée par le serveur pendant la période d'enregistrement indiquée. Ce calcul est effectué à partir des fichiers log eux-mêmes ce qui peut présenter quelques erreurs de calcul.

° Remarque : un KByte équivaut à 1024 Bytes, et 1Mb=1000 Kbytes. Pour atteindre les 3Gb de trafic auquel vous avez droit, il faut donc 3'000'000 Kbytes.

Navigateur

Un navigateur est un logiciel client qui affiche des pages au format HTML et permet l'accès au World Wide Web. La visualisation d'un site est dépendante du navigateur avec lequel il est parcouru.

Cette rubrique permet de connaître les navigateurs qui ont été utilisés par les visiteurs du site et de savoir quel est le navigateur le plus fréquemment utilisé, et donc d'optimiser la visualisation de votre site pour ce type de navigateur.

Pages

Cette valeur est un sous-ensemble des Fichiers. Le critère de comptabilisation est le type du fichier (déterminé par son extension). Typiquement c'est le nombre d'accès aux fichiers avec comme extension : .htm, .html, .cgi, .pl. Pour déterminer la fréquentation de son site web on pourra choisir cette valeur. Ce comptage n'inclut pas les comptages de l'affichage de fichiers graphiques et de sons. On peut appeler ce compteur : pages vues.

Tout document HTML, ou qui génère une page HTML, est considéré comme une page. Par contre, les documents inclus dans un document HTML comme une image ou un son ne sont pas considérés comme des pages.

Sites ou hôtes

C'est le nombre de machines clientes différentes qui sont venues visiter le site web. Ces chiffres peuvent être utilisés pour déterminer la Fréquentation du site. Ceci NE SIGNIFIE PAS le nombre d'utilisateurs individuels uniques (une vraie personne) qui ont visité, ce qui est impossible à déterminer en utilisant seulement des fichiers logs et le protocole HTTP (cependant, ce chiffre pourrait être très près de cette réalité)

° Attention : Il n'est pas possible de comptabiliser le nombre exact de personnes physiques qui ont visité le site.

° Par exemple, si la connexion utilisateur se fait par serveur proxy, des visiteurs distincts ne seront comptabilisés qu'une fois.

° Remarque : Le «Pays» est déterminé par l'extension du site ayant généré la requête. Mais cette information n'est pas sûre car elle est à la résolution de l'IP par un nom de domaine, et l'extension de la connexion n'est pas forcément liée au pays (une connexion bluewin.ch est correctement liée à la Suisse, mais une connexion aol.com n'est pas liée à un pays). En effet, l'utilisateur d'un nom de domaine .COM peut résider aux Etats-Unis ou ailleurs.

De plus, un grand pourcentage de connexions est indiqué comme Irrésolue / Inconnu car la résolution du nom de domaine n'est pas toujours possible à effectuer.

URL, Pages d'entrée et Pages de sortie

Les personnes connectées directement sur le site à partir de la page de news par exemple ont soit mémorisé l'URL de cette page lors d'une précédente visite, soit eu accès au site à partir d'un autre site web qui les a renvoyé directement sur la page de news, ou encore ont trouvé cette page en utilisant un moteur de recherche.

Ces informations peuvent éventuellement vous permettre de remanier l'enchaînement de vos pages de manière à mieux adapter votre site à la consultation par les internautes.

° Remarque : l'URL racine "/" indique la page d'accueil du site.

° Remarque : Les « Chaînes de recherche » (Search Strings) sont obtenues en examinant les chaînes des référents et en regardant les modèles connus des différents moteurs de recherche. Le plus souvent les chaînes de recherche correspondent aux mots clés que vos visiteurs ont utilisés pour vous trouver en utilisant un moteur de recherche.

Visites

À chaque fois qu'une requête est faite au serveur à partir d'une adresse IP (un client) donnée, la durée de temps depuis une requête antérieure de cette adresse (si elle a lieu) est calculée. Si la différence de temps est plus grande que la valeur pré-configurée du "temps mort de visite" (ou s'il n'y a eu aucune requête auparavant), alors cela est considéré comme une "nouvelle visite", et ce total est incrémenté (à la fois pour le site et l'adresse IP).

La valeur par défaut du "temps mort de visite" est de 30 minutes. Ainsi, si un internaute visite le site à 13H00 et qu'il y revient ensuite à 15H00, alors deux visites seront enregistrées dans le calcul, tandis que s'il revenait à 13h20, une seule visite serait enregistrée.

Liste détaillée et signification des codes retournés par les Serveurs HTTP

Voici la liste complète de ces codes ainsi que leur signification.

200	OK
201	Créé. Suite à une commande POST, la réponse consiste en un URI par lequel le document créé est identifié
202	Accepté. La requête a été acceptée pour traitement mais ce dernier n'a pas été finalisé
203	Information partielle. En réponse à une commande GET, ###
204	Pas de réponse. Le serveur a bien reçu la requête mais n'a pas d'informations à renvoyer. C'est généralement le cas lorsqu'un script doit être exécuté sans que le document ne soit modifié
301	Déplacé. Les données demandées ont reçu un nouvel URI, cette modification étant permanente
302	Trouvé. Les données réclamées ont été trouvées bien qu'elles soient définies sous un autre URL
303	Méthode. Suggère que le programme client utilise une autre adresse URL mais également une autre méthode
304	Non modifié. Le document a été trouvé mais n'a pas été modifié depuis la date précisée dans le champ "If-Modified-Since"
400	Requête incorrecte. La requête est de syntaxe incorrecte ou impossible à exécuter
401	Non autorisé
402	Paiement requis. Le client doit recommencer sa requête en précisant dans son en-tête la mention "ChargeTo"
403	Interdit
404	Document non trouvé
500	Erreur. Le serveur a rencontré une condition non prévue qui l'empêche d'exécuter la requête
501	Non implémenté. Le serveur ne supporte pas la fonction demandée
502	Service temporairement surchargé. Le serveur ne peut provisoirement répondre à la demande suite à un trafic important
503	Gateway timeout. Apparaît lorsque le serveur, à son tour, fait appel à un service extérieur qui ne répond pas dans les délais impartis