Google Dorks

By Haye Grégory

Sommaire

- Petit focus sur le moteur de recherche Google
- Google dorks, Google Hacking ?
- Opérateurs de recherche
- Google Hacking : GHDB
- Sources

Petit focus sur le moteur de recherche Google

Stats

Mécanismes & Algorithmes …

Quelques stats

- Parts de Marché de Google 2021 : +65% devant Safari (18%)
- Requêtes sur Google :
  - 130 000 milliards de pages y sont indexées (2017)
  - 20 milliards de sites visités par google (Crawlées) chaque jour
  - 80 000 requêtes chaque seconde, soit 6.9 milliards par jour
  - 15 % des requêtes sont de nouvelles requêtes (+- 500 millions par jour !)
  - + 110 millions de Go de données sont stockées sur leurs serveurs
  - +90% du trafic des recherches en France provient de Google (2019)

Mécanismes, Algorithmes

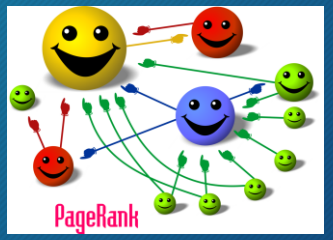
3 Mécanismes

- Explorer le web
  - Crawler / Spider
  - Robots.txt
- Indexer les pages
  - DB (url, balises html, liens etc.)
- Classer les résultats
  - Interpréter la requête de l'utilisateur
  - Identifier les pages
  - Classer les pages et les afficher par pertinence

Repose sur des algorithmes :

- PageRank

$$PR(p_i) = \frac{1-d}{N} + d \sum_{p_j \in M(p_i)} \frac{PR(p_j)}{L(p_j)}$$



  - Inventé par Larry Page (1997)
- RankBrain
  - Machine learning (2016)
  - Evolution vers un moteur de réponse
- Et bien d'autres animaux …

Google Docks , Google Hacking ?

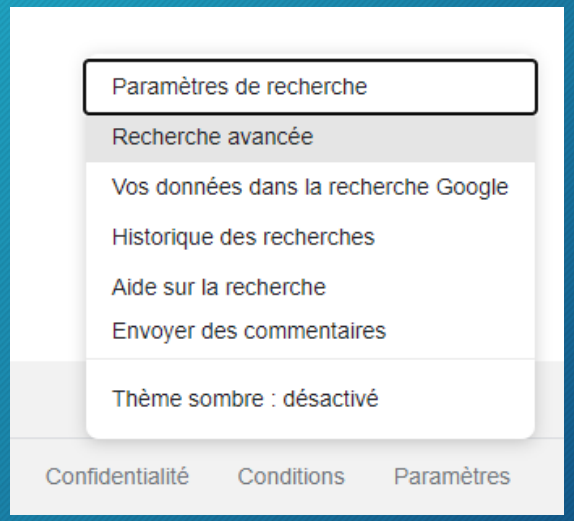
Quesaco ?

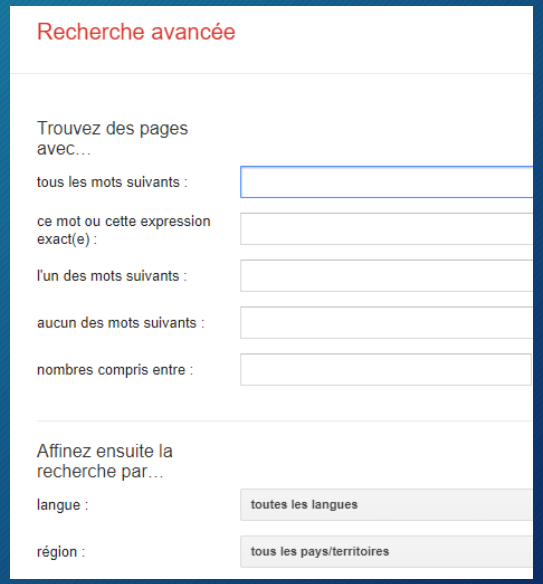
Que peut-on trouver avec les Google Dorks ?

Dorking défensif

Quésaco ?

- Méthodes de recherche avancées
- Moteur de recherche de Google (mais aussi sur d'autres)
- S'appuie sur le renseignement en sources ouvertes (OSINT)





Que peut-on trouver avec les Google Dorks ?

- Equipements non protégés exposés sur Internet :
  switchs, caméras, routeurs, imprimantes, etc.
- Fichiers sensibles :
  liste du personnel, liste d'utilisateurs et mots de passe, etc.
- Fichiers correspondant à des listes de prix (pricelist)
- Pages d'authentification sur des applications Web :
  espace d'administration d'un site, PhpMyAdmin, etc.
- Serveurs exposés sur Internet et mal configurés voire pas configurés : page par défaut d'Apache
- Etc...

Les intentions comptent et la notion d’éthique est importante !

Dorking défensif

- Vérification des failles de sécurité pour un service en ligne
  Site web, serveur FTP, …
- Recherche d'informations sensibles (ou sur une personne avec sa permission) exposées involontairement sur le site
  - PDF, XLS, DOC, …
  - Adresses, numéro de téléphone
  - Adresses IP serveurs
  - …

- Cependant attention. Si vous recherchez votre nom ou votre adresse, puis, disons, votre numéro de sécurité sociale, vous donnez en principe cette information à quiconque lance l'outil de recherche.

Opérateurs de recherche

Quelques opérateurs

Quelques opérateurs

Opérateur	Fonction	Exemple
« »	Forcer une recherche de correspondance exacte	« Larry Page »
OR |	Renvoie des résultats liés à X ou Y, ou aux deux.	PageRank OR BrainRank
AND	Renvoie que les résultats liés à la fois à X et à Y.	PageRank AND BrainRank
-	Exclure un terme ou une expression.	jobs -apple
*	Caractère générique et correspondra à n'importe quel mot ou expression.	steve * apple
( )	Regroupez plusieurs termes ou opérateurs de recherche	(ipad OR iphone) apple
$ €	Rechercher des prix. Fonctionne également pour l'euro (€)	ipad $329
Define:	Un dictionnaire intégré à Google	Define:Quésaco
Cache:	Renvoie la version en cache la plus récente d'une page Web indexée	cache:apple.com
filetype: ext:	Restreindre les résultats à ceux d'un certain type de fichier.	apple filetype:pdf / apple ext:pdf

Quelques opérateurs

Opérateur	Fonction	Exemple
Site:	Limitez les résultats à ceux d'un site Web spécifique.	site:apple.com
Related:	Rechercher des sites liés à un domaine donné.	related:apple.com
Intitle:	Trouvez des pages avec un certain mot (ou des mots) dans le titre.	intitle:apple
Allintitle:	Idem mais tous les mots spécifiés dans la balise de titre.	allintitle:apple iphone
Inurl:	Trouvez des pages avec un certain mot (ou des mots) dans l'URL.	inurl:apple
Allinurl:	Idem mais tous les mots spécifiés dans l'URL.	allinurl:apple iphone
Intext:	Trouvez des pages contenant un certain mot (ou des mots) quelque part dans le contenu.	intext:apple
Allintext:	Idem mais tous les mots spécifiés quelque part sur la page.	allintext:apple iphone
Around(x)	Trouvez des pages contenant deux mots ou expressions à X mots l'un de l'autre	apple AROUND(4) iphone
Weather:	Recherchez la météo d'un lieu spécifique.	weather:New-York

Quelques opérateurs

Opérateur	Fonction	Exemple
Stocks:	Consultez les informations sur les actions.	stocks:Atos be
Map:	Afficher les résultats de la carte pour une recherche de localisation.	map:silicon valley
Movie:	Rechercher des informations sur un film spécifique.	movie:whoami
In	Convertir une unité en une autre.	$329 in EUR
Source:	Rechercher des résultats d'actualités provenant d'une certaine source dans Google Actualités.	apple source:the_verge
_	Pas exactement un opérateur de recherche, mais agit comme un joker pour Google Autocomplete.	apple CEO _ jobs

Il en existe d'autres plus aléatoires (.. , inanchor:, etc.) ou même abandonnés par google (+, ~, link, etc.)

Google Hacking : GHDB

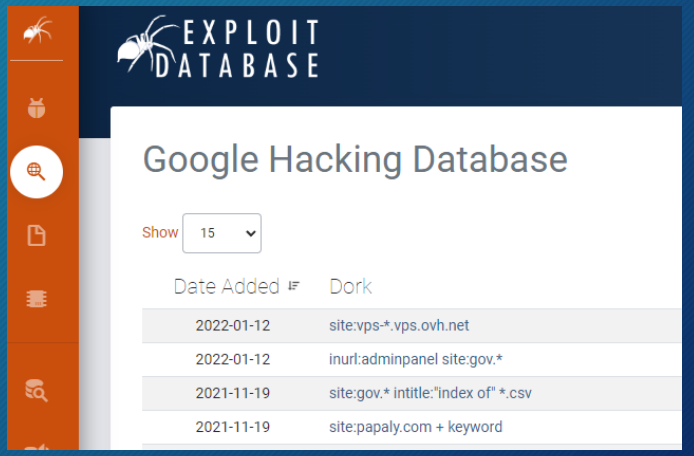
Exploit Database

D’autres outils

Exploit Database

Avec ces opérateurs nous pouvons créer nos propres requêtes (Notre seule limite est l'orée de notre imagination - Dixit Kuoni)

- Ceci dit, sachez que le site exploit-db.com contient une section nommée "Google Hacking Database"(GHDB):
  - + 6 500 requêtes Google Dorks différentes !
  - régulièrement maj par la communauté du site.
  - Un véritable moteur de recherche pour Google Dorks !



D’autres outils

- https://github.com/opsdisk/pagodo
- https://github.com/blueudp/DorkMe
- https://dorksearch.com/
- Vos outils ?

Sources

https://www.blogdumoderateur.com/chiffres-google/

https://www.it-connect.fr/google-dorks-google-hacking-exploiter-toute-la-puissance-de-google/

https://smartkeyword.io/seo-algorithme-google-machine-learning-focus-rankbrain/

https://audreytips.com/google-comment-ca-marche-explorer-indexer-classer/

https://www.netoffensive.blog/referencement-naturel/premier-sur-google/fonctionnement/algorithmes-google/rankbrain/

https://www.netoffensive.blog/referencement-naturel/premier-sur-google/fonctionnement/algorithmes-google/pagerank/

https://redback-optimisation.fr/la-liste-des-algorithmes-de-google/

https://ahrefs.com/blog/google-advanced-search-operators/

https://www.exploit-db.com/google-hacking-database

Il y a toujours de la lumière au bout de chaque tunnel.

☺ Merci de votre attention ☺