

تبدأ العملية بإرسال وزيارة الروبوتات Googlebot الخاصة بجوجل والبرامج الذكية الاصطناعي للمواقع الجديدة من خلال التنقل من صفحة إلى صفحة أخرى، وهي برامج كمبيوتر تُسمى "عناكب الويب" (Web Crawlers)، تقوم هذه الروبوتات بزيارة وقراءة صفحات الويب واكتشاف الروابط الموجودة عليها وجميع ملفات الموقع بشكل تلقائي من خلال ملف اسمه robots.txt. sitemap، ويتم تخزين هذه الروابط في فهرس ضخم (Index) من المعلومات عن صفحات الويب هدفها فهرسة جميع الصفحات الخاصة بالموقع. يقوم السيرفر بطلب من قوئل برفع الملفات على خوادمها من خلال إرسال البوت الخاصة بجوجل لعمل زحف وقراءة الملفات بهدف فهرسة جميع صفحات الموقع ومعرفة ماذا يوجد داخل كل صفحة من محتوى ومعرفة URL الخاص بها. ممكن ان يتم الزحف والارشفة من خلال جوجل دون الطلب او بشكل عشوائي وتلقائيا تتم الفهرسة بفترات متباعدة وهذا لا ينصح به، وتحدث هذه العملية من خلال ظهور موقعك بروابط في مواقع أخرى وظهرت للعناكب فوراً يتم ارشفتها بسبب وصولها وزحف العناكب لها، وفي حال حدثت هذه العملية لا تتم علمية الفهرسة والتحديث بسهولة لموقعك لذلك انت بحاجة لاعادة الزحف والفهرسة مع تحديد موعد الفهرسة مرة أخرى سواء يومي او اسبوعي بإبلاغ عناكب جوجل بتعديلات أو تحديثات الموقع باستخدام الأداة المجانية الخاصة بالأرشفة والتابعة لجوجل اسمها جوجل سيرش كونسول Google Search Console. من خلال تحديد خرائط الموقع Sitemap الخاص بكل موقع يتم انشاء وبناء ملف يسمى robots.txt يستخدم لأخبار الروبوت بفهرسة الصفحات المحددة التي يسمح لها بالفهرسة وإظهارها في النتائج والتي لا يسمح لها بالفهرسة مثل: صفحة التحكم وغيرها من الصفحات الغير مرغوب بها لإظهارها في النتائج، يتم رفع على جوجل سيرش كونسول وربطه بالموقع. ملف robots.txt هو ملف نصي بسيط يُستخدم لإخبار محركات البحث بأي صفحات ويب على موقعك الإلكتروني يجب زيارتها وفهرستها (crawl and index) وأي صفحات لا يجب زيارتها وفهرستها، ويتكون ملف robots.txt. المستخدم: (User-agent) تُحدد هذه المجموعة أي روبوتات ويب (web crawlers) يُسمح لها بزيارة موقعك الإلكتروني. يمكنك حظر جميع روبوتات الويب أو حظر روبوتات ويب معينة (مثل روبوتات Googlebot أو Bingbot). السماح/الحظر: (Allow/Disallow) تُحدد هذه المجموعة أي صفحات ويب على موقعك الإلكتروني (URLs) يُسمح لروبوتات الويب بزيارة وفهرستها (crawl and index) وأي صفحات لا يجب زيارتها وفهرستها. كيف يرتبط الزحف وقراءة موقعك ملف robots.txt. تبدأ محركات البحث بزيارة موقعك الإلكتروني بشكل عشوائي على أحد المواقع وفي حال ظهر موقعك بروابط لمواقع أخرى يتم ارشفته من خلاله وصول العناكب (crawling) لقراءة ملف robots.txt. يُخبرهم ملف robots.txt بأي صفحات ويب (URLs) يُسمح لهم بزيارة الصفحة وقراءة محتواها. إذا لم يُسمح لروبوت ويب بزيارة صفحة ويب (URL) فإنها لن تُضاف إلى فهرس محرك البحث. بمجرد أن يُسمح لروبوت ويب بزيارة صفحة ويب (URL) فإنه سيقراً محتوى الصفحة. يستخدم روبوت ويب هذا المحتوى لفهم موضوع الصفحة وملاءمتها للاستعلامات التي يبحث عنها المستخدمون. في حال تم عمل زحف وفهرسة لموقعك تلقائية يجب متابعة تحديث موقعك. يتم تخزين هذه المعلومات في فهرس محرك البحث. لا يضمن ملف robots.txt منع ظهور موقعك الإلكتروني في نتائج البحث. إذا تم ربط موقعك الإلكتروني بمواقع ويب أخرى (backlinks) فقد تظهر صفحات موقعك الإلكتروني في نتائج البحث حتى لو لم تتم فهرستها بواسطة روبوت ويب. من المهم تحديث ملف robots.txt بانتظام (regularly) إذا قمت بإجراء تغييرات على موقعك الإلكتروني من خلال سيرش كونسول Google search console ومهم كمتخصص سيو أنك تبلغ بأي تحديث في موقعك. على سبيل المثال، إذا قمت بإضافة صفحات جديدة أو إزالة صفحات قديمة، فيجب عليك تحديث ملف robots.txt لإعلام روبوتات الويب بالتغييرات. هناك العديد من الأدوات المتاحة لمساعدتك في إنشاء ملف robots.txt. يمكنك العثور على مزيد من المعلومات حول ملف robots.txt على مواقع ويب محركات البحث.