

# Web crawler : est-ce que tout est permis ?



ASIE

## ■ INTRODUCTION

Dans cette ère de « l'Economie de la donnée » (« data economy »), les entreprises sont amenées à exploiter des volumes importants de données qui, une fois analysées, peuvent avoir un impact significatif sur leur productivité et leur rentabilité. Ces dernières années, par exemple, la technologie financière (FinTech) s'est considérablement développée sous l'effet du Big data mais a aussi fait l'objet de discussions animées en matière de conformité. L'utilisation de données massives impliquent de toujours garder à l'esprit la question suivante : quelle est la source des données et de quelle manière ont-elles été collectées ?

A cet égard, les robots d'exploration (en anglais « Web crawlers ») ne doivent pas être négligés en tant qu'outil puissant de collecte de données.

Le Web crawler, également connu sous le nom de « Web spider » ou « Web robot », est un programme ou un script qui capture automatiquement les données publiées sur Internet en fonction de certaines règles. La principale fonction d'un Web crawler est de parcourir automatiquement le réseau, de recueillir des informations et de construire un index.

Le premier Web crawler est né aux Etats-Unis en 1993. Aujourd'hui, les Web crawlers les plus connus sont Googlebot de Google, Baiduspider de Baidu, Applebot d'Apple. En bref, sans Web crawler, pas de moteurs de recherche !

Développé depuis plus de 30 ans, le rôle du Web crawler est aujourd'hui essentiel dans la bataille sur les données. Les entreprises peuvent l'utiliser directement pour collecter des données sur Internet (images, vidéos, textes, etc.), ou recourir à des prestataires qui fournissent à leurs clients des produits ou services se basant sur les données collectées par des Web crawlers. Dans les faits, une utilisation vertueuse des Web crawlers peut être un vrai moteur de croissance. En revanche, un usage non contrôlé des Web crawlers peut entraîner la responsabilité civile et même pénale des entreprises.

## ■ LE "CODE DE CONDUITE" DES WEB CRAWLERS

Amasser des données, est la raison d'être du Web crawler. Les Web crawlers peuvent-ils pour autant explorer librement n'importe quelle donnée ? Si vous répondez oui à cette question, les cas d'usage suivants pourraient vous faire changer d'avis.

Afin de restreindre l'exploration sans limite des Web crawlers, un ingénieur-réseaux hollandais a proposé en 1994, un « Code de Conduite » pour les Web crawlers. Selon ce principe, lorsque le Web crawler atteint un site web, le propriétaire du site peut afficher le « Code de Conduite » du site afin d'encadrer l'activité du Web crawler (quelles sont les données qui peuvent être explorées et celles qui ne peuvent pas l'être). Ainsi, avant d'explorer les données, les Web crawlers doivent lire attentivement le « Code de Conduite » du site et s'y conformer. Si le Code de Conduite n'autorise aucun accès aux données du site, le Web crawler doit passer son chemin. Dans la pratique, le « Code de Conduite » mentionné ci-dessus s'appelle « Protocole d'Exclusion des Robots / Protocole des Robots » (plus communément connu sous le nom « Robots.txt »), et constitue l'une des stratégies anti-exploration les plus courantes. Toutefois, le Protocole d'Exclusion des Robots n'est pas une mesure technique contraignante, mais un simple « Avis » affiché sur la « porte » du site web. Par conséquent, ce Protocole ne peut que donner des instructions aux Web crawlers, mais il ne peut pas empêcher les activités d'exploration malveillantes. Il n'est donc pas surprenant que des problèmes liés à l'exploration illégale de sites web se posent.

Cas d'usage n° 1 : Le premier cas s'est produit aux Etats-Unis en 1999. Le robot d'exploration du site Web de Bidder Edge a violé le Protocole d'Exclusion des Robots défini par le site web d'eBay et a exploré toutes sortes de données commerciales du site eBay. Le tribunal a interdit à Bidder Edge de collecter des données sur le site d'eBay par des moyens techniques (y compris des Web crawlers) sans l'autorisation d'eBay.

Cas d'usage n° 2 : Le deuxième cas s'est déroulé en 2017 et concernait LinkedIn et HiQ. LinkedIn compte plus de 600 millions d'utilisateurs dans le monde

*La Newsletter a pour but de donner un aperçu des développements juridiques récents. Son contenu n'exprime pas un avis juridique et ne saurait se substituer à une consultation juridique.*

et propose à ses utilisateurs différents niveaux de stratégie de protection de la vie privée. Les utilisateurs peuvent choisir d'afficher leurs données personnelles au public/aux contacts/au réseau de contacts. Lorsque les utilisateurs choisissent d'être complètement « publics », leurs données personnelles sont accessibles à des utilisateurs qui ne figurent pas sur LinkedIn et aux moteurs de recherche comme Google. HiQ est une entreprise qui fournit un service de suivi des ressources humaines et de Talent mapping pour les employeurs, qui se base principalement sur des données collectées sur LinkedIn. Après avoir été mise en demeure de cesser ses agissements, HiQ a attaqué LinkedIn en justice afin d'obtenir une injonction pour empêcher LinkedIn de bloquer l'accès aux données publiques de ses utilisateurs. HiQ a été soutenue par le tribunal. En effet, le tribunal a estimé que l'interdiction d'accès aux données de profils publics par LinkedIn n'était pas favorable à un marché sain et avait pour conséquence de fausser le jeu de la libre concurrence.

Il est évident que les deux cas ci-dessus se terminent différemment et la différence de traitement réside dans la manière dont la technologie a été utilisée, dans la relation de concurrence entre les entreprises en question et surtout dans le stade de développement de l'économie de la donnée. Ainsi, l'utilisation conforme d'un Web crawler est une question traitée au cas par cas, et la situation peut varier selon le pays.

## ■ LES WEB CRAWLERS EN CHINE

Selon l'Article 127 du Code Civil de la RPC, « si la loi comporte des dispositions sur la protection des données et la propriété virtuelle du réseau, ces dispositions prévalent ». Cependant, il n'existe pas en Chine, de réglementation directement liée à la protection des données en tant qu'actif immatériel d'une entreprise. Faut-il comprendre pour autant que l'utilisation des Web crawlers se fait en dehors de tout cadre réglementaire en Chine?

En fait, la Loi chinoise Contre la Concurrence Déloyale a déjà été appliquée dans des affaires impliquant des Web crawlers et certaines entreprises ont été reconnues coupables de concurrence déloyale (voir notamment le cas d'usage n° 3 ci-dessous).

Dans la pratique, les juridictions ont tendance à protéger davantage la partie dont les données sont explorées par des Web crawlers. Cela s'explique par le fait que le Big Data, l'intelligence artificielle, l'Internet des objets et autres technologies sont encore au stade initial de développement en Chine : il y a une volonté de protéger les investissements effectués par les acteurs dans ce domaine.

Cas d'usage n° 3 : En 2016, Baidu (moteur de recherche Chinois) a été poursuivi par Dazhong Dianping (le plus important site d'avis et conseils de consommateurs) pour avoir illégalement crawlé les informations des commerçants et les commentaires des consommateurs, en violation du Protocole d'exclusion des robots du site. L'intégralité des informations ainsi crawlées étaient visibles directement sur la page de Baidu, qui se substituait ainsi à Dazhong Dianping et permettait de générer plus de trafic à son bénéfice. Le tribunal a jugé que la conduite de Baidu représentait un acte de concurrence déloyale et qu'il devait verser des dommages et intérêts de 3,23 millions de RMB (environ 400 000 Euros) à Dazhong Dianping.

Il convient de noter que les conséquences de l'utilisation inappropriée d'un Web crawler peuvent être plus importantes que la seule indemnisation du dommage causé. L'usage illégal d'un Web crawler peut constituer une infraction : collecte illégale des données de systèmes informatiques, destruction de système informatique, etc. (voir le cas d'usage n° 4 ci-dessous).

De plus, quand les informations collectées par un Web crawler concernent des œuvres protégées par le droit d'auteur ou portent sur des données personnelles, cela peut être constitutif d'une violation des droits d'auteur ou un manquement au respect des règles sur la protection des données personnelles. Dans ce cas, les auteurs de ces actes seraient également tenus pour responsables et sanctionnés. Il ne fait aucun doute que de tels agissements auront un impact négatif sur le crédit social de l'entreprise elle-même et sur le crédit social des dirigeants concernés.

Cas d'usage n° 4 : En 2017, la première procédure pénale chinoise au sujet d'un Web crawler, s'est déroulée entre la start-up Xiutao (plateforme vidéo pour le commerce électronique) et Toutiao (plateforme vidéo du groupe Bytedance). Il est intéressant de noter que dans ce cas, deux des fondateurs de Xiutao étaient d'anciens salariés de Toutiao : l'un d'eux était chargé des produits et des opérations tandis que l'autre était responsable du service technique de Toutiao. Xiutao a collecté des données vidéo de Toutiao par l'intermédiaire d'un agent utilisateur et en évitant les mesures techniques anti-extraction mises en place par Toutiao. Xiutao et plusieurs de ses dirigeants ont été poursuivis pour avoir commis le délit d'obtention illégale de données de systèmes

*La Newsletter a pour but de donner un aperçu des développements juridiques récents. Son contenu n'exprime pas un avis juridique et ne saurait se substituer à une consultation juridique.*



informatiques. Dans cette affaire, non seulement Xiutao a été condamné à payer des dommages et intérêts, mais les dirigeants concernés de Xiutao ont été condamnés à une peine de prison ferme.

## ■ L'AVENIR DES WEB CRAWLERS

Actuellement, le Web crawler est l'outil de base utilisé par les entreprises du Big Data. Une utilisation conforme et raisonnable de cette technologie peut aider les entreprises à réduire leurs coûts et accroître leur productivité. Cependant, de nombreuses entreprises et individus pensent encore que les «données publiques» peuvent être exploitées librement, sans contrainte ou limite. En outre, quand ils utilisent un Web crawler ou reçoivent des informations obtenues par un Web crawler et fournies par un tiers (comme des rapports sur les tendances du marché, etc.), ils sont susceptibles d'ignorer les questions de conformité liées à l'utilisation des Web crawlers.

Par conséquent, afin de réduire les risques juridiques, nous suggérons aux entreprises de prendre sérieusement en compte au moins les points suivants avant d'utiliser un Web crawler et/ou d'autres technologies pour collecter des données :

- Respecter le Protocole des Robots ;
- Respecter les dispositions relatives à la protection des données personnelles et notamment le principe de minimisation dans la collecte des données
- Respecter les droits d'auteur ;
- Analyser la relation de concurrence entre les parties ;
- S'assurer du respect du principe de nécessité dans la collecte et l'usage des informations collectées



*Pour toute information complémentaire, merci de contacter :*

**ZHANG Beibei**  
Associate - Shanghai Office  
[beibeizhang@dsavocats.com](mailto:beibeizhang@dsavocats.com)

**Isabelle DOYON**  
Lawyer - Shanghai Office  
[doyon@dsavocats.com](mailto:doyon@dsavocats.com)