

Avec Memex, la Darpa scrute les tréfonds de l'Internet

Deep web ou darkweb : tout comme les icebergs, Internet dispose d'une face visible mais aussi d'une partie immergée. Cette dernière, tout comme les énormes glaçons flottants, est beaucoup plus importante que son équivalent émergé, mais difficile à quantifier et surtout à observer. L'agence américaine pour les projets de recherche avancée de défense, **Darpa**, s'est dotée d'un moteur de recherche un peu particulier : [Memex](#). Ce nom, contraction de **Memory et Index**, fait aussi référence au nom d'un terminal dans un article « As We May Think », écrit par Vannevar Bush, directeur de l'OSRD (équivalent du Darpa pendant la seconde guerre mondiale) en 1945.

Les équipes de l'émission 60 minutes de [CBS](#) ont discuté avec le chef du projet Memex, Chris White. En préambule, le chercheur explique que les moteurs traditionnels de recherche comme Google ou Bing ne fournissent aux utilisateurs que **5% du contenu qui existe sur le Web**. L'indexation auprès du grand public se fait à travers des algorithmes liés au classement des sites et au ciblage publicitaire ; « *Internet est beaucoup, beaucoup plus grand que ce que les gens pensent* », ajoute le spécialiste.

Un Google sous stéroïdes

D'où l'idée de travailler sur « *un Google sous stéroïdes* » pour aller explorer le banc, l'arrière-banc et les tréfonds du web. Fruit d'un travail de plus d'un an et mobilisant 17 équipes d'ingénieurs issues de plusieurs sociétés sous contrat avec l'agence militaire, Memex a pour ambition de bâtir une meilleure cartographie des contenus du Web. Pour atteindre cet objectif, le moteur de recherche recense des millions de pages qui ne sont pas référencées par les moteurs classiques, mais aussi des pages qui sont présentes sur le darkweb où l'anonymat est de rigueur.

Dans le viseur de Memex, on trouve **le réseau Tor** que de multiples sites utilisent et pour lesquels il est obligatoire de télécharger le client ad hoc. Or il existe peu de moteur de recherche pour indexer le contenu de ces sites cachés. Chris White évoque quelques initiatives comme Grams, mais qui ciblent des thématiques particulières (drogues, médicaments, armes). Memex apporte des estimations plus affinées sur le nombre de sites cachés qui utilisent Tor, « *avant nous avions une estimation d'un millier de sites, maintenant nous sommes capables d'observer jusqu'à 40 000 sites opérant à un moment donné* », souligne Chris White. L'agence y ajoute une dose d'analytique pour trouver les relations entre des sites qui, en apparence, n'ont rien de commun.

Une aide pour les enquêtes judiciaires

Mais derrière ce moteur de recherche, il s'agit aussi d'**aider les autorités judiciaires** à traquer, à pister certains comportements délictueux comme la traite de personnes (prostitution, esclavage, etc) sur le darkweb. Il existe plusieurs axes de recherche : forum, chat, sites cachés, publicité, etc. Avec Memex, il est possible de **trouver des correspondances** (localisation, adresse IP, terminaux utilisés, etc) sur des annonces de prostitutions avec les mêmes femmes qui se déplacent dans

plusieurs localités aux États-Unis (dans la démonstration réalisée pour le reportage télé). Il s'agit d'un simple exemple, mais il peut s'appliquer à d'autres thématiques comme **les armes ou la drogue**. La Darpa annonce que 8 partenaires dont les noms sont gardés secrets travaillent sur des prototypes intégrant la puissance de Memex. Certains y verront un outil de surveillance supplémentaire...

A lire aussi :

[La DARPA publie un catalogue de projets open source](#)

[La DARPA planche sur de l'électronique destructible par déclenchement](#)

Crédit Photo : Releon8211-Shutterstock