
Documents sauvegardés

Mardi 19 mai 2020 à 17 h 10

1 document

Par Bibliothèque Municipale de Lyon

Sommaire

Documents sauvegardés • 1 document

-
- | | | |
|----------------------|---|----------|
| La Matinale du Monde | <p>15 mai 2020</p> <p>A la BNF, les archivistes du Web sauvegardent l'Internet français du confinement</p> <p>Depuis le 17 mars, date à laquelle la France a été placée en confinement pour lutter contre la pandémie de coronavirus, les archivistes du Web français de la bibliothèque nationale de ...</p> | 3 |
|----------------------|---|----------|



Nom de la source

La Matinale du Monde

Type de source

Presse • Presse Web

Périodicité

Quotidien

Couverture géographique

Internationale

Provenance

Paris, Ile-de-France, France

Vendredi 15 mai 2020

La Matinale du Monde • 819 mots

A la BNF, les archivistes du Web sauvegardent l'Internet français du confinement

Par Pauline Croquet

Depuis le 17 mars, date à laquelle la France a été placée en confinement pour lutter contre la pandémie de coronavirus, les archivistes du Web français de la bibliothèque nationale de France (BNF) compilent une somme de documents en ligne en rapport avec le Covid-19, et l'impact qu'il a eu dans la vie quotidienne des internautes français. A terme, cette collecte permettra aux chercheurs d'enquêter et de retrouver des archives Internet qui auraient pu disparaître.

A côté de l'immense sauvegarde annuelle effectuée automatiquement à l'automne par un « crawler » appelé Heritrix, l'équipe de dix bibliothécaires et experts informaticiens dédiés au dépôt légal du Web à la BNF procède également à des collectes ciblées pour des événements ponctuels (élections, événements sportifs, temps forts imprévus...). « Les dernières opérations de collecte d'urgence que nous avons menées concernaient par exemple les attentats de 2015, le mouvement #metoo, mais aussi celui des "gilets jaunes", l'incendie de Notre-Dame », énumère Benoît Tuleu, directeur du département du dépôt légal à la BNF.

Recueillir toutes les facettes de la pandémie



GUILLAUME MURAT / BNF

Le Web français archivé par le département du dépôt légal est consultable par les chercheurs sur le site de la BNF.

« Fin janvier, l'équipe permanente avait commencé à s'intéresser au coronavirus en France dans sa veille habituelle notamment via les premiers hashtags francophones qui apparaissaient comme #JeNeSuisPasUnVirus qui dénonçait la stigmatisation de la communauté asiatique, explique Benoît Tuleu, mais le fait que la population soit confinée du jour au lendemain tout en restant très connectée nous paraissait inédit et nécessairement intéressant à archiver pour les études dans le futur. »

Expérience d'autant plus inédite que le personnel de la BNF aux manettes de l'opération était lui-même confiné et en télétravail. Mais depuis un mois, l'équipe permanente du dépôt légal numérique a été renforcée par la veille de 50 bibliothécaires supplémentaires en interne à la BNF. Un réseau de correspondants de 26 bibliothèques territo-

© 2020 SA Le Monde. Tous droits réservés. Le présent document est protégé par les lois et conventions internationales sur le droit d'auteur et son utilisation est régie par ces lois et conventions.



Certificat émis le 19 mai 2020 à BM-LYON à des fins de visualisation personnelle et temporaire.

news-20200515-LMT-6039704

riales et services d'archives a été activé, pour un affinage local des documents collectés

« Notre collecte Covid-19 suit l'évolution et l'impact global de la pandémie sur le Web français. On se concentre sur toutes les facettes de la crise sanitaire : médical, scientifique, culturel, vie quotidienne, rapport au corps, questions morales, sociabilisation, etc. Dans un souci de représentativité on recense aussi bien des comptes officiels que de personnes, des blogs, des journaux de **confinement**, des contenus publics sur les réseaux sociaux mais aussi des vidéos sur YouTube », liste Benoît Tuleu.

Les adresses des sites Web ou contenus à sauvegarder sont pointées manuellement par les bibliothécaires auprès du robot qui va les scanner à des rythmes différents selon la pertinence : plusieurs fois par jour sur un réseau social par exemple; une fois par mois pour un site moins alimenté; une indexation quotidienne pour les sites de presse. Pour l'heure, ce sont 2 000 adresses URL liées au Covid-19 que le robot a collectées.

Entre 8 et 10 téraoctets

Environ 40 % de ce contenu sur le coronavirus provient des réseaux sociaux, estiment les bibliothécaires. « Difficile d'estimer aujourd'hui la taille de cette collecte qui n'est pas encore terminée. Mais si on extrapole sur ce qui a déjà été réalisé, le volume la collection sur le coronavirus devrait se situer entre 8 et 10 téraoctets de données, selon la durée de l'épidémie. Pour comparaison, l'ensemble des collectes annuelles a un volume de 150 téraoctets », précise Benoît Tuleu.

A la différence du dépôt légal que la

BNF opère pour les livres mais aussi pour d'autres supports comme le jeu vidéo, les archivistes du Web ne sont pas exhaustifs — ce serait tout bonnement impossible de conserver l'entièreté de la Toile — et n'attendent pas que les propriétaires des sites leur fassent parvenir une copie. « Ce sont nos équipes qui vont moissonner les sites avec en tête un souci de représentativité, sans jugement sur la qualité ni d'a priori, dans une démarche encyclopédique », indique Benoît Tuleu. Une mission posée par la loi Dadvsi de 2006 mais que les bibliothécaires avaient commencé à expérimenter plusieurs années auparavant; et complétée par une acquisition d'archives du Net français entre 1996 et 2000, de précieux « *incunables du Web* » comme ils les appellent.

Aujourd'hui, ce sont en tout deux pétaoctets (soit deux millions de gigaoctets) d'Internet sauvegardé que la **BNF** a rassemblés sur ses serveurs à raison d'à peu près 2,6 milliards d'adresses URL scannées chaque année. Ils sont consultables par les chercheurs sur place dans le 13^e arrondissement parisien et dans un réseau de bibliothèques partenaires sur le territoire français.

Outre les champs de recherche à venir, ce travail de curation dédié au coronavirus stocké sur les serveurs de la **BNF** va être partagé avec le consortium International Internet Preservation Consortium (IIPC) pour contribuer à l'archivage international Novel Coronavirus (2019-nCoV) outbreak, lancé en février 2020 en association avec Internet Archive. Une mémoire collective de la crise sanitaire à laquelle participe une trentaine de bibliothèques dans le monde.

Cet article est paru dans La Matinale

du Monde

[/pixels/article/2020/05/15/a-la-bnf-les-archivistes-du-web-sauvegardent-l-internet-francais-du-confinement_6039704_4408996.html](https://pixels/article/2020/05/15/a-la-bnf-les-archivistes-du-web-sauvegardent-l-internet-francais-du-confinement_6039704_4408996.html)