

Une extension qui meta-press.es à ta disposition

La veille sur la presse en ligne est laborieuse et exigeante, mais une extension pour Firefox peut la rendre plus légère, rapide et efficace... et plus éthique que Google News.

Nous rencontrons aujourd'hui Simon Descarpentries pour lui poser des questions sur le module Meta-Press.es qu'il a créé.

Bonjour Simon, pourrais-tu nous dire par quels chemins tu es arrivé à Meta-press.es...



Commençons par le début : je suis né en 1984 comme la FSF et comme elle, je préfère mes logiciels avec de l'éthique en plus d'un code source accessible (pour reprendre cette belle formule de Pouhiou). J'ai découvert GNU+Linux en l'an 2000, ai adhéré à l'April en 2002, terminé mes études en 2007 et travaillé pour Framasoft

de 2009 à 2011. J'ai ensuite rejoint Sopinspace¹ jusqu'à sa mise en sommeil 4 ans plus tard et mon entrée dans Acoeuro.com².

Très tôt j'ai décidé de ne pas pousser le monde dans la mauvaise direction le jour en essayant de compenser les dégâts bénévolement la nuit. Je me suis donc efforcé de gagner ma vie en faisant directement ce qui me semble éthique et via Acoeuro.com j'ai une grande liberté de choix dans mes clients, qui ne sont que des associations, clubs sportifs ou collectivités locales³.

Mais ce n'est pas tout d'en vivre, j'ai aussi toujours eu envie de contribuer au Logiciel Libre en retour. J'ai commencé par des traductions (avec la formidable équipe Framalang), puis me suis attelé à la comptabilité de FDN.fr (pendant 5 ans), j'ai fait un peu de JavaScript pour lancer un widget de campagne de LQDN.fr ou aider à éplucher les 5000 réponses d'une consultation de la Commission Européenne... Mais ça me démangeait toujours.

Je me suis donc également occupé de la revue de presse de La Quadrature du Net pendant 5 ans, et c'est là que m'est venue l'envie de développer une alternative à Google News, afin de libérer l'association de sa dépendance envers un acteur qu'elle critique à juste titre le reste du temps. À la faveur d'un inter-contrats en 2017, j'ai repris mes prototypes précédents de méta-moteur de recherche pour la presse et j'ai exploré sérieusement cette piste.

..tu en arrives ainsi à Meta-Press.es ?

Voilà, j'ai réservé un nom de domaine dès que j'ai eu une preuve de concept fonctionnelle.

Alors c'est quoi exactement ?

Meta-Press.es est un moteur de recherche pour la presse sous forme d'une extension pour Firefox.

Directement depuis notre navigateur, il interroge un grand nombre de journaux⁴. L'extension récupère les derniers résultats de chaque journal et permet de..

- trier ces résultats,
- mener notre recherche dedans,
- sélectionner ceux qui nous intéressent,
- exporter cette sélection suivant plusieurs formats (JSON, RSS ou ATOM, bientôt CSV aussi).

Et comment ça marche au juste ?

Eh bien il faut bien sûr installer l'extension depuis sa page officielle.

Tu ouvres ensuite l'onglet de l'extension en cliquant sur



l'icône tu saisis les termes de ta recherche, tu précises les sources dans lesquelles chercher : journaux, radios, agrégateurs de publications scientifiques... par défaut tu choisis celles qui sont dans ta langue, et tu lances la recherche.



Meta-Press.es va alors interroger les sources choisies et afficher les résultats.

The screenshot shows the Meta-Press.es website interface. At the top, there is a navigation bar with links for 'Index', 'Import', 'JSON', 'RSS', 'ATOM', 'A propos', and a question mark. Below the navigation bar is a large logo for 'META-PRESS.ES' with the tagline 'DECENTRALIZED PRESS META-SEARCH ENGINE'. The logo features a stylized net with a star and a black star. Below the logo, there are several news snippets with icons indicating their type (e.g., CHECKNEWS, VIDEO, Boris Johnson, Black Friday). A search bar is present with the text 'reconnaissance faciale' and a green play button. Below the search bar, there are filters for 'Select sources by themes', 'Lang:', 'Country:', 'Periodicity:', and 'Tags:'. The 'Lang:' filter is set to 'fr (24)'. The 'Country:' filter is set to 'be (6)' and 'fr (30)'. The 'Periodicity:' filter is set to 'daily (94)'. The 'Tags:' filter is set to 'https (95)'. Below the filters, there are several news snippets with titles and brief descriptions. On the right side, there is a table showing the number of results for each source. The table has a 'Total:' row at the bottom with a value of 1259. Below the table, there is a 'Filter results' section with a dropdown menu and three radio buttons for 'Last 2 wks', 'Last 2 mon.', and 'Last 6 mon.'.

Source	Count
Le Figaro	661
Mediapart.fr	320
Libération	197
Euronews (fr)	35
L'Humanité	31
Metro Time (be)	15
B2 Bruxelles2	0
Total:	1259

Quelles différences avec un agrégateur RSS ou une poche-kangourou comme Wallabag ?

Contrairement à un agrégateur de flux RSS, Meta-Press.es donne

accès aux contenus qui existaient avant qu'on s'abonne aux flux, puisque Meta-Press.es utilise la fonctionnalité de recherche des journaux. L'extension s'emploie à propager la requête de l'utilisateur auprès de chaque source pour agréger tous les résultats et les trier dans l'ordre chronologique. Google News ne fait pas beaucoup plus en apparence, or ça, n'importe quel ordinateur peut le faire.

Ensuite, Meta-Press.es intègre déjà un catalogue de sources connues (principalement des journaux, mais aussi des radios ou des agrégateurs de publications scientifiques), et est directement capable de chercher dans toutes ces sources, alors qu'un agrégateur de flux RSS doit être configuré flux par flux.

Actuellement, la base contient un peu plus de 100 sources (de 38 pays et en 21 langues), dont déjà 10 % ont été ajoutées par des contributeurs. Ces sources sont organisées par un système d'étiquettes pour les thèmes abordés, la langue ou d'autres critères techniques. Ce système permet d'accueillir toutes les contributions et l'utilisateur choisira ensuite dans quoi il veut chercher.

Apparemment c'est surtout pour faire de la veille sur la presse, est-ce que ça peut intéresser tout le monde ou est-ce un truc de « niche » pour un nombre limité de personnes qui peuvent y trouver des avantages ?

L'extension a été développée avec le cas d'usage de la revue de presse de la Quadrature du Net en tête. On est toujours plus efficace en grattant soi-même là où ça démange.

Mais les journalistes auxquels j'ai présenté Meta-Press.es se sont également montrés enthousiastes, car l'outil renvoie toujours les mêmes résultats quand on fait les mêmes recherches (même si on change d'ordinateur ou de connexion internet). Ce n'est pas le cas quand ils utilisent Google News, car l'entreprise traque leur comportement (historique

des recherches, articles consultés) pour renvoyer ensuite des résultats de recherche « personnalisés » (donc différents d'une fois sur l'autre), et surtout, pour vendre aux annonceurs de la publicité ciblée.

Et puis un Mastonaute a récemment trouvé un autre moyen de se servir de Meta-Press.es :



Au-delà des recherches, ce que permet Meta-Press.es, c'est d'exporter les résultats d'une recherche dans un fichier. On peut ainsi :

- archiver les résultats ;
- les reprendre plus tard (même hors-connexion) ;
- les envoyer à un ami.

On peut également sélectionner les résultats que l'on souhaite exporter. Une sélection exportée au format RSS peut ensuite facilement être ajoutée au flux RSS de la revue de presse d'une association (comme c'est le cas pour LQDN).

Cela épargne les deux-tiers du travail dans ce domaine, qui consistait sinon à copier chaque information (titre, date, source, extrait...) de la page du journal à l'outil générant le flux de la revue de presse.

D'ailleurs, pendant mes années de revue de presse à LQDN, plusieurs associations amies nous ont demandé quels outils on utilisait de notre côté. On a répondu à chaque fois qu'on

utilisait plusieurs outils faits main (la nuit par l'un des cofondateurs) emboîtés les uns dans les autres sans documentation technique et que l'essentiel du travail restait fait à la main. C'était frustrant pour tout le monde.

Aujourd'hui, l'outil est là, revenez les amis, nous pouvons désormais tous nous partager Meta-Press.es !



Logo de l'extension par
Elisa de Guerra-Castro

____*____*____

Bon, pendant que les plus impatients sont déjà en train de tester l'extension et de s'en servir, nous avons d'autres lecteurs et lectrices un peu plus techniques qui veulent savoir comment ça se passe dans les coulisses, et pour commencer avec quelles briques tu as travaillé...

Techniquement, c'est fait en JavaScript moderne (ECMAScript6/7) avec tous les meilleurs exemples trouvés sur StackOverflow, developer.mozilla.org et surtout les API récentes comme : *async/await*, `<script type= »module »>`, *fetch*, *crypto*, *domParser*, *XPathEvaluator*...

Je suppose que certain·es vont vouloir ajouter des « sources »...

Le processus de contribution des sources a été simplifié au maximum.

D'ailleurs, la première contribution à Meta-Press.es en a détourné l'usage pour y intégrer une dizaine de sources de publications scientifiques. C'est un cas auquel je n'avais pas pensé en développant Meta-Press.es, mais la contribution s'est facilement intégrée au reste du projet, on a juste eu à y mettre l'étiquette de thème : « science ».

- Si la source propose des résultats de recherche sous la forme d'un **flux RSS**, il suffit en gros de préciser le nom de la source et l'adresse du flux (en remplaçant les termes de recherche par : {}) dans le champ d'exemple de source des paramètres de l'extension. La plupart des journaux qui utilisent le moteur WordPress fournissent ce flux (sûrement sans même le savoir), ce qui représente un tiers des sources de Meta-Press.es à l'international. En revanche en France, beaucoup de journaux utilisent SPIP comme moteur de site web, or SPIP n'intègre pas cette fonctionnalité par défaut... Une mise à jour serait très bienvenue ! En proportion, les flux RSS représentent donc environ 30% des sources actuelles, contre 1% de flux ATOM (la bataille est moins serrée qu'entre vim et emacs).
- Si la source ne propose **pas de résultats en RSS**, l'exercice peut se résumer à rassembler une URL et 4 sélecteurs CSS. Toutefois, il faut donc parler HTML et CSS couramment, et bien souvent RegEx aussi pour arriver à ses fins. J'ai listé des documentations synthétiques dans le README.adoc du projet sur Framagit.

Une source est donc décrite par un objet JSON dans lequel on détaille comment accéder à chaque information dans la page de résultats (par des sélecteurs CSS donc pour pointer les éléments) et éventuellement en ajoutant un retraitement du texte obtenu par motif de remplacement en expression rationnelle. Pour s'exercer, l'ajout d'une source peut se

faire directement depuis les paramètres de l'extension (où vous trouverez des conseils et les sources fournies en exemples). Si vous avez un résultat fonctionnel, vous pouvez vous contenter de me l'envoyer simplement par courriel, je m'occuperai de l'intégrer au projet.

J'avais testé en 2013 plusieurs solutions pour voir ce qui était le plus rapide dans le rapatriement et l'analyse des pages web listant les résultats de recherche de chaque source. Entre autres, les pages de résultat sont analysées par Firefox dans ce qu'on appelle un fragment de HTML. Ce dernier n'est pas complètement interprété par le navigateur web (pas de rendu graphique). Entre autres, les images et les feuilles de style de la page ne sont pas chargées. Il n'y en a pas besoin. En se contentant du HTML les choses se passent bien plus rapidement que s'il fallait charger les mégaoctets d'images et de traqueurs des journaux.

En puisant les résultats dans des flux RSS, le traitement devrait aller encore plus vite car la structure XML d'un flux RSS est minimaliste. Mais les serveurs web priorisent mal ces requêtes, alors les flux RSS mettent facilement plus de 10 secondes à arriver.

Une autre question épineuse, liée au JavaScript est celle de l'analyse des dates en format non américain. De nombreuses bibliothèques de fonction existent pour parer aux déficiences de la norme, mais elles sont généralement volumineuses et lentes. Je propose ma propre solution dans ce domaine, avec la fonction JavaScript `month_nb` qui se contente de transformer un nom de mois en son numéro, mais sait le faire pour 69 langues et n'a même pas besoin de connaître la langue du mois à convertir. J'en ai parlé plus longuement sur le site de Meta-Press.es notamment pour détailler l'aspect minimaliste et ré-utilisable de mon approche : il fallait là aussi « faire rentrer le monde dans un fichier JSON » mais ça c'est bon, c'est fait.

Mais pourquoi avoir choisi une extension pour Firefox plutôt qu'une appli pour Android ou une appli *standalone* à installer sur son ordinateur... ?

Je faisais tenir mes premiers prototypes dans un unique fichier HTML. Je trouvais ça élégant d'avoir tout dans un seul fichier : le code, l'interface graphique, les données... Et puis un fichier HTML c'est facile à distribuer (par clé USB, en pièce jointe d'un courriel, directement sur le web...). Toutefois, comme je l'ai expliqué dans le billet « Motivations » du blog du projet, une contrainte technique empêchait ce modèle de fonctionner pour Meta-Press.es : on ne peut pas accéder au contenu d'une *iframe* depuis le JavaScript d'une simple page web.

Et puis j'ai compris qu'avec une extension pour Firefox la contrainte pouvait être levée. J'ai donc tout naturellement continué mon travail dans cette direction, en m'appuyant sur les technologies que je manipule au quotidien : le Web.

Avec un peu de recul, je considère que c'était une excellente idée. Firefox est probablement l'analyseur de HTML le plus rapide au monde, en cours de ré-écriture, par morceaux, en Rust. Piloter cette fusée via un langage de script se révèle à la fois plaisant et efficace.

Si j'avais voulu faire une application à part, j'aurais probablement utilisé le langage Python (dont je préfère la syntaxe, surtout édité avec vim et des tabulations !), mais j'aurais forcément eu à manipuler un analyseur de page web moins rapide et probablement moins à jour.

Ensuite, en tant qu'extension de Firefox, Meta-Press.es est aussi utilisable avec le navigateur web Tor, qui est taillé pour la protection de votre vie privée et installable en quelques clics sur n'importe quel ordinateur et quasiment n'importe que système d'exploitation.

Le navigateur Tor a été inventé en grande partie pour lire la presse en ligne sans être suivi, ni laisser de traces. Les deux font donc la paire. Avec le navigateur Tor les journaux

ne savent pas qui vous êtes, et avec Meta-Press.es vous n'avez plus besoin de Google pour les trouver. Retour au modèle du bon vieux journal lu dans le fauteuil du salon, sans autres conséquences, ni à court, ni à long terme.

En ce qui concerne Android, l'extension fonctionne parfaitement une fois installée sur Firefox pour Android (ou la version IceCatMobile en provenance de la logithèque libre pour Android : F-Droid.org).

Mozilla offre l'avantage de fournir l'infrastructure de distribution du programme et un référencement (l'extension est facile à retrouver via le moteur de recherche d'addons.mozilla.org avec les mots-clés « meta presse »). Mozilla gère les mise à jour, des retours utilisateurs rapides ou complets via les commentaires, la notation par étoiles et même une porte de collecte de dons pour soutenir le projet – qui fonctionne très bien ;-).

Si l'on ajoute la documentation et les recommandations suite à l'analyse du code (automatique mais aussi effectuée par des humains), c'est une plateforme très accueillante.

Dans l'actualité récente les éditeurs de presse en ligne français étaient en conflit avec Google et son moteur de recherche. Est-ce que de nouvelles contraintes légales ne vont pas impacter Meta-Press.es ?

Oui, j'ai suivi ce feuilleton, et non ça ne devrait avoir de conséquence pour Meta-Press.es.

Pour reprendre un peu le sujet, tout se joue autour de la directive européenne sur les éditeurs de presse en ligne, que les élites du gouvernement se sont empressées de transposer en droit français, pour l'exemple et avec de grandes annonces.

Cet épisode a donné lieu en septembre à de savoureux échanges entre Google et les éditeurs. Le fond du problème était que les éditeurs, déjà sous perfusion de l'État, ont cru qu'ils pourraient taxer Google aussi (en améliorant la rente de leur

situation, plutôt qu'en s'adaptant à un monde qui change), au moins pour un montant proportionnel à l'extrait d'article que Google republie chez lui, à côté de ses publicités, et dont de plus en plus de lecteurs se contentent (comme je l'ai détaillé dans ce commentaire sur LinuxFR.org).

Ça m'a fait bizarre, mais c'est Google que j'ai trouvé de bonne foi pour le coup : aucune raison de payer la rançon. Le géant américain a d'ailleurs simplement répliqué en retirant les extraits visés, en publiant des stats sur la faible consultation des résultats de Google News sans extrait, et en indiquant que pour un retour aux affaires il suffisait de préciser son accord via un fichier hébergé par chaque journal (une directive du fameux robot.txt).

En deux semaines la moitié des éditeurs avaient autorisé Google à reprendre gratuitement les extraits, au bout d'un mois tous avaient rejoint le rang. Tout ce travail législatif international pour en arriver là : un communiqué de presse du moteur de recherche et des redditions sans condition de la presse.

Aujourd'hui c'est facile à dire, mais je pense que les éditeurs n'ont pas pris le bon chemin... Au lieu d'essayer de jouer au plus malin et de perdre magistralement⁵, ils devraient chercher à s'émanciper de cet intermédiaire qui valorise sa pub avec leurs contenus. Un moyen de se débarrasser de cet intermédiaire, ce serait de développer eux-mêmes un Meta-Press.es, rien ne l'empêche techniquement. Après, j'ai quelques années d'avance, mais rien ne les empêche non plus de me soutenir.

Je me suis logiquement fait quelques sueurs froides, inquiet de voir bouger l'horizon juridique d'un projet sur lequel je me suis attelé depuis plusieurs années. Mais je vais pouvoir laisser les extraits de résultats de recherche dans Meta-Press.es, car cet outil n'entre pas dans le périmètre d'application de la loi, qui ne vise que les plateformes

commerciales, ce que n'est pas Meta-Press.es. De plus, Meta-Press.es ne publie rien, tout se passe entre le navigateur d'un internaute et les journaux, pas d'intermédiaire.

Pas d'intermédiaire, mais plein d'idées pour continuer le développement de l'outil ?

Ça oui ! À commencer par l'indispensable mise en place d'un cadriciel (*framework*) de test automatisé des sources, pour tenir toute la collection à jour en détectant celles dont la présentation des résultats a changé et doit être revue.

Ensuite, j'ai déjà évoqué l'ajout d'un format d'export CSV ou la présentation de l'extension sur écran de téléphone, mais l'outil pourrait par exemple également être internationalisé pour en diffuser plus largement l'usage.

Une grande idée serait d'implémenter un test de rapidité de réponse des sources, pour ne retenir que les sources qui répondent rapidement chez vous.

La possibilité de récupérer plus que les 10 derniers résultats de chaque source est également sur les rails, et en fait malheureusement, la TODO-list du projet ne fait que s'agrandir au fur et à mesure que je travaille à la réduire...

Comment vois-tu la suite pour Meta-presse.es ?

Meta-Press.es n'est pas une grande menace pour Google, mais c'est une alternative techniquement viable.

Il faut maintenant faire l'inventaire des journaux du monde⁶ et mettre cet index en commun dans le dépôt des sources de Meta-Press.es. Je n'y arriverai pas seul, mais je suis bien déterminé à faire cette part de dé-Google-isation de l'internet⁷ et à la faire bien, dans la plus pure tradition Unix (une chose à la fois, mais bien faite).

Cela fait déjà des années que je travaille sur Meta-Press.es et je porterai ce projet le plus loin possible. Avec moi une contribution n'est jamais gaspillée. Alors je compte sur vous

pour m'aider à indexer la presse en ligne.

Je vous encourage à bidouiller votre source préférée et à me l'envoyer si elle fonctionne ou si vous avez besoin d'un coup de pouce pour terminer. Indiquez-moi par courriel les sources à flux RSS que vous avez trouvées car elles sont très rapides à intégrer, et normalement stables dans le temps⁸.

Cet inventaire, réalisé pour un projet libre et fait dans un format standard (JSON) sera réutilisable à volonté. C'est une autre garantie qu'aucune contribution ne sera perdue.

D'autres malices dans ta boîte à projets ?

Avec les connaissances acquises en développement d'extension pour Firefox, il y a d'autres problèmes auxquels j'aimerais proposer des solutions... Je pense par exemple au paiement en ligne sur le Web. C'est parce qu'il n'y a pas de moyen simple de payer en un clic que la plupart des éditeurs de contenus s'empressent de grever leurs œuvres de publicité, parce que ça, au moins, ça rapporte, et sans trop d'efforts.

Une solution pourrait être proposée sous la forme d'une extension de Firefox. Une extension qui lirait le contenu des liens affublés du protocole payto: (comme il existe déjà le mailto:), ouvrirait une fenêtre de sélection de banque, proposerait de vous loguer sur votre compte via le site officiel de votre banque, et vous avancerait en lecture rapide jusqu'à la validation d'un virement bancaire, pour le destinataire précisé dans le lien payto:, pour le montant, le libellé et la devise précisée.

Dans l'idéal, les banques proposeraient une interface pour faire ça facilement, mais elles ne le font pas, et on n'en a pas forcément besoin pour que ça marche, il suffit d'arpenter leur interface web comme on le fait pour les résultats de recherche des journaux avec Meta-Press.es.

Coupler cette idée avec les virements rapides que les banques

sont en train de concéder pour faire face au Bitcoin, et voilà, le Web serait réparé...

S'il y a des financeurs que ça intéresse, moi je sais faire...

____*____*____

- Aller sur le site officiel de Mozilla pour Télécharger l'extension Meta-press.es
- Pour en savoir plus sur l'extension et les nouveautés de la récente version 1.2
- Le code de l'extension sur son dépôt Framagit

____*____*____