

Se connecter

Proposer un contenu

Identifiant

Identifiant

Mot de passe

Mot de passe

☐ Connexion automatique

Se connecter

Pas de compte ? S'inscrire...

# Journal : Démerdifier le web avec Unmerdify et Offpunk

Posté par **ploum** (site web personnel, Mastodon) le 01 décembre 2025 à 17:20.  
Licence CC By-SA.

Étiquettes : **offpunk, smallweb, smolnet, gemini, merdification**



Vous le savez certainement, beaucoup de choses autour de nous se merdifient :

<https://ploum.net/2023-06-15-merdification.html>

Et le Web ne fait pas exception. C'est même plutôt devenu la norme : chaque site web semble vouloir enfouir son contenu dans des centaines de liens, de kilooctets de javascript, de pubs, d'espionnage.

On peut évidemment se concentrer sur le "small web/smolnet" :  
<https://tcrouzet.com/2025/11/16/smolweb-smolnet/>

Voire, pour les plus aventureux, sur le protocole Gemini:  
<https://ploum.net/gemini-le-protocole-du-slow-web/index.html>

Mais c'est dommage de se passer de plein de contenus vraiment pertinents et pourtant enfouis dans leur merdier.

C'est ici qu'intervient l'utilitaire "unmerdify", créé par Vincent Jousse.  
<https://codeberg.org/vjousse/unmerdify/>

Unmerdify se base sur une série de règles maintenues par la communauté et, au départ, créées pour instapaper. Chaque domaine possède des transformations XSLT pour extraire le contenu correct en supprimant la "merde", ou, pour le moins, ce qui n'est pas très utile/central.

<https://github.com/fivefilters/ftr-site-config>

Vous pouvez y contribuer, il n'y a pas encore de règles pour linuxfr par exemple.

Tout seul, unmerdify n'est pas très utile.

C'est pourquoi, en coopération avec Vincent, j'ai intégré Unmerdify dans Offpunk.

Offpunk est un navigateur en ligne de commande:  
<https://offpunk.net/>

Pour se lancer avec offpunk, il suffit de l'installer, de lancer la commande "offpunk" puis "tutorial". (les retours et contributions sont bienvenues)

Jusqu'à présent, pour tenter d'extraire du contenu d'une page web, offpunk utilisait la librairie "readability". Readability utilise un algorithme statistique. Cela veut dire qu'il considère que, par défaut, le contenu intéressant d'un page est une fraction de la page. Et que s'il y a beaucoup de liens, ce sont sans doute des pubs.

Paradoxalement, un site bien fait avec le contenu au centre va donc être agressivement réduit par Readability. C'est le cas de mes billets de blog où, par exemple, des billets "en vrac" de Tristan Nitot sur le Standblog. De plus, readability n'est pas vraiment maintenu.

Le but est donc de remplacer Readability par Unmerdify dans Offpunk. La fonctionnalité est dans TRUNK depuis aujourd'hui.

<https://lists.sr.ht/~lioploum/offpunk-devel/%3C176459371683.7.10565078656826518034.1039550217@ploum.eu%3E>

Pour l'activer, il faut d'abord avoir une copie locale du ftr-site-config:

\$ git clone <https://github.com/fivefilters/ftr-site-config>

puis le dire dans offpunk:

```
$ offpunk
ON> set ftr_site_config /path/to/my/ftr-site-config
```

Et voilà.

Pour ne pas avoir à retaper cette commande à chaque démarrage, on peut bien entendu la mettre dans son offpunkrc

Cerise sur le gateau, il est également possible d'utiliser unmerdify tout seul en ligne de commande, tout comme les autres composants de Offpunk:

- netcache (un curl/wget qui garde les données en cache de manière permanente)
- ansicat (une commande cat qui fait un joli rendu dans le terminal des fichiers HTML/RSS/Atom/Gemtext/Gophermap et même les images)
- opnk (un "open" universel qui fait le rendu automatiquement ou trouve le meilleur logiciel pour afficher un contenu particulier)

Bref, si vous avez envie de tester offpunk/unmerdify, je suis preneur de vos feedbacks.

Et n'hésitez pas à poster sur la mailing-liste utilisateurs, que ce soit des questions, des retours ou des idées.

<https://lists.sr.ht/~lioploum/offpunk-users>

(3 commentaires).

Markdown

EPUB

## # Interface graphique

Posté par **ploum** (site web personnel, Mastodon) le 01 décembre 2025 à 17:22.

Évalué à 4 (+2/-0).

J'oubliais de préciser qu'il est tout à fait possible d'imaginer une interface graphique basée sur offpunk pour permettre aux gens n'aimant pas la ligne de commande de naviguer sur un web démerdifié. Une de mes étudiantes à réalisé un prototype.



--

Mes livres CC By-SA : <https://ploum.net/livres.html>

Répondre

## # Pour chaque site ?

Posté par **Xanatos** le 01 décembre 2025 à 17:43. Évalué à 3 (+1/-0).

Hello,  
Lu vite, si j'ai bien compris créer un fichier de config, site après site.  
Au final on fait le boulot d'indexation des crawlers des moteurs de recherche d'antan.



Quitte à me faire taper dessus: combattre le mal par le mal ; des agents LLM qui iraient faire ce boulot à notre place (crawl liste> grep structure; echo fichier)

Cela dit j'aime bien l'idée de la couche d'abstraction sur le contenu pur, ça a peut être une dénomination ? Après tout des pâquerettes peuvent naître du fumier.  
Mes 2 cents.

Répondre

### [^] # Re: Pour chaque site ?

Posté par **ploum** (site web personnel, Mastodon) le 01 décembre 2025 à 18:30. Évalué à 3 (+1/-0).

Oui, c'est pour chaque site. Mais s'il n'y a pas de règle, offpunk retombe par défaut sur Readability.



Il y a déjà plus de 2000 sites sur ftr-site-config et c'est surtout pour les plus merdifiés que c'est intéressant.

--

Mes livres CC By-SA : <https://ploum.net/livres.html>

Répondre

Envoyer un commentaire

Suivre le flux des commentaires

**Note** : les commentaires appartiennent à celles et ceux qui les ont postés. Nous n'en sommes pas responsables.

Revenir en haut de page

Derniers commentaires

- A320
- Re: Zenodo ?
- Re: Pour chaque site ?
- Re: Le changelog
- Merki
- Pour chaque site ?
- Re: Jour 1
- Re: Jour 1
- Re: Le changelog
- Interface graphique
- Re: Leaderboard
- Re: Équilibre délicat

Étiquettes (tags)

populaires

- intelligence\_artificielle
- merdification
- grands\_modèles\_de\_lan...
- entretien
- administration\_française
- bulle
- linux
- jeu\_vidéo
- android
- tour\_des\_gull
- adieu\_windows
- capitalisme

Sites amis

- April
- Agenda du Libre
- Framasoft
- Éditions D-BookeR
- Éditions Eyrolles
- Éditions Diamond
- Éditions ENI
- La Quadrature du Net
- Lea-Linux
- En Vente Libre
- Grafik Plus
- Open Source Initiative

À propos de LinuxFr.org

- Mentions légales
- Faire un don
- L'équipe de LinuxFr.org
- Informations sur le site
- Aide / Foire aux questi...
- Suivi des suggestions et...
- Règles de modération
- Statistiques
- API pour le développe...
- Code source du site
- Plan du site