

Capitole du Libre 2023, et mon exposé sur la censure de l'Internet

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 22 novembre 2023

<https://www.bortzmeyer.org/capitole-du-libre-2023.html>

Les 18 et 19 novembre 2023, c'était Capitole du Libre <<https://capitoledulibre.org/>> à Toulouse. Plein d'exposés intéressants et de rencontres passionnantes autour du thème du logiciel libre.

C'est gros, le Capitole du Libre : 1 200 personnes sont venues, il y avait 118 orateurices, 94 conférences, 18 ateliers et 31 stands. Comme il y a eu davantage de propositions que de créneaux, certaines ont été refusées. 50 kg de pop-corn ont été légués par le DevFest <<https://devfesttoulouse.fr/>> dont 20 ont été consommés. Mais il y avait aussi des pains au chocolat :

Personnellement, j'ai fait un exposé sur la censure sur l'Internet. Les supports au format PDF sont (et leur source en). La vidéo a été faite mais n'est pas encore publiée. (Amusant calendrier, cet exposé sur la censure a été fait deux jours avant la publication du RFC 9505¹.)

Ironie de la situation, le réseau de l'ENSEEIH, l'école qui héberge gratuitement le Capitole du Libre, met en œuvre une des techniques de censure présentées. Si on tente de se connecter à OnlyFans, on récupère un RST ("ReSeT") TCP, probablement suite à l'examen du SNI dans l'ouverture de la session TLS. Vu par tcpdump, voici ce que ça donne (le dernier paquet, le RST, n'est pas légitime) :

```
13:19:35.603931 IP 172.22.223.35.40874 > 54.211.0.120.443: Flags [S], seq 2413541568, win 64240, options [mss 14
13:19:35.744066 IP 54.211.0.120.443 > 172.22.223.35.40874: Flags [S.], seq 3076610656, ack 2413541569, win 26847
13:19:35.744173 IP 172.22.223.35.40874 > 54.211.0.120.443: Flags [.], ack 1, win 502, options [nop,nop,TS val 46
13:19:35.830075 IP 172.22.223.35.40874 > 54.211.0.120.443: Flags [P.], seq 1:518, ack 1, win 502, options [nop,n
13:19:35.833848 IP 54.211.0.120.443 > 172.22.223.35.40874: Flags [R.], seq 1, ack 518, win 502, length 0
```

1. Pour voir le RFC de numéro NNN, <https://www.ietf.org/rfc/rfcNNN.txt>, par exemple <https://www.ietf.org/rfc/rfc9505.txt>

Et ne dites pas d'utiliser Tor, les nœuds d'entrée publics sont bloqués.

Maintenant, les bons exposés (dans l'ordre arbitraire que j'ai décidé tout seul avec mes préférés en premier). Vincent Privat a présenté un système pour récupérer automatiquement les photos que publie la NASA (avec leurs métadonnées) et les mettre sur Wikimedia Commons. C'est légal puisque la NASA, comme toutes les agences fédérales étatsuniennes, met tout ce qu'elle publie dans le domaine public. Mais c'est techniquement difficile. Déjà, s'il y a parfois une API pour lister les images à récupérer, parfois, il faut plutôt "*scraper*" l'HTML. Et parfois il faut deviner l'URL en testant plusieurs variantes. D'autre part, il faut éliminer les doublons (chacun peut écrire dans Wikimedia Commons et ces images sont souvent importées). Son outil récupère toutes les métadonnées, les vérifie et convertit vers ce qu'attend Wikimedia Commons. Comme toujours avec l'"*open data*", il y a des bavures amusantes, par exemple la date de prise de vue indiquée est parfois dans le futur, il y a du mojibake dans les textes, des licences incohérentes (entre le texte et les EXIF), des « descriptions » qui sont en fait de la publicité, etc (tous des problèmes classiques de l'"*open data*"). D'autre part, beaucoup d'outils de Wikimedia sont conçus pour le texte, car la cible est Wikipédia. Utiliser les outils et machines de Wikimedia pour traiter des images de beaucoup de giga-octets (par exemple pour le dédoublement) explosait les quotas. Enfin, pour dédoublement, le programme calcule un "*perceptual hashing*", pas un simple et traditionnel condensat.

À noter que les autres agences spatiales ne mettent rien sous licence libre donc la vision de l'espace est dominée par celle des États-Unis, les seuls à avoir compris la force du "*soft power*".

Julian Vanden Broeck a fait un très bon bilan de son changement du système de messagerie instantanée interne dans son entreprise (passage à Matrix). Comment ça marche en vrai? Quel accompagnement pour les utilisatrices? Quelle acceptation? Entre utilisatrices qui râlaient contre l'ancien système, celles qui râlaient contre le nouveau, celles qui disaient que Microsoft Teams, c'était mieux, et celles qui critiquaient la couleur des boutons, l'acceptation n'était pas évidente. La technique, c'est facile, ce sont les humains qui sont compliqués. « On a désactivé le chiffrement systématique car c'était trop frustrant pour les utilisateurs, notamment en cas de perte de clés. » Je confirme que c'est aussi mon expérience avec Matrix, loin des promesses d'un système parfait. Et encore, l'entreprise de l'orateur n'utilise que le client Element, les problèmes sont pires quand on a plusieurs clients (de ce point de vue, Matrix ne marque aucun progrès par rapport à XMPP, où l'incompatibilité des clients a toujours été une plaie). Au passage, si vous voulez essayer, je suis joignable à `@bortzmeyer:matrix.underworld.fr`.

On s'est bien amusés avec l'exposé « Modèles d'organisation ouverts dans les entreprises du logiciel libre » car les deux orateurs ne se sont pas présentés, sans explication. Thomas Petazzoni a alors lancé l'idée que, puisque le sujet était les organisations ouvertes, non hiérarchiques, etc, on n'avait pas besoin d'orateurs et on pouvait débattre entre nous. Ça a très bien pris et il y a eu une discussion intéressante, par exemple avec les gens d'une association, l'Échappée Belle <<https://github.com/lechappeebelle>>, qui est en fait une coopérative de "*freelances*", où chaque salarié décide de son salaire. Évidemment, ils insistent sur le fait que ça ne peut sans doute marcher que dans les petites organisations (ils sont six), où les gens se connaissent et ont des valeurs communes.

Et Pierre-Yves Lapersonne a posé la question du choix des bibliothèques logicielles quand on développe un logiciel libre. Le ou la programmeuse ne peut pas tout faire, il faut donc dépendre de bibliothèques existantes. Vous, je ne sais pas mais moi, quand je commence un projet, j'ai toujours du mal à identifier les bonnes bibliothèques dont mon projet va dépendre. Vous utilisez quoi, comme critères? Quand il y a le choix, c'est compliqué. Comment choisir, et, par exemple, détecter celles qui ne sont plus maintenues (donc où les bogues ne seront plus corrigés), celles gérées par une boîte qui va subitement changer la licence (il y a eu deux-trois cas spectaculaires ces dernières années, comme MongoDB), etc. Parmi les critères possibles (mais aucun n'est parfait et l'auteur recommande de se méfier des critères quantitatifs simples) :

- la licence utilisée, bien sûr,
- le langage de programmation utilisé,
- les dépendances de cette bibliothèque,
- le nombre de tickets, s'il est faible, c'est peut-être qu'il n'y a pas beaucoup d'utilisateurs, si beaucoup sont fermés, cela peut être bon signe (bogues corrigées rapidement) ou mauvais signe (mainteneur qui ferme tout de suite sans corriger),
- la date du dernier ticket fermé peut être un bon indicateur pour savoir si le projet vit toujours,
- nombre de "commits" (pas très pertinent car il dépend du style de travail des programmeurs, certain-es font dix "commits" dans la matinée, d'autres un par semaine (pensez à `git rebase...`)),
- date du dernier "commit" pour détecter les codes abandonnés (ou bien tellement parfaits qu'ils n'ont pas besoin de changements...),
- l'activité du mainteneur; est-ce qu'il est actif (voir son profil sur la forge publique utilisée) avec le langage du projet (un projet Go créé par quelqu'un qui ne fait que du Rust aujourd'hui risque de ne pas faire l'objet de beaucoup d'attention),
- l'existence de tests, et la couverture du code par la suite de tests (pour avoir une base saine si on va devoir modifier le logiciel).

On a aussi discuté d'UI et d'UX au Capitole du Libre. Nojhan a parlé de LiquidPrompt, un des nombreux mécanismes permettant d'avoir des invites très riches pour la ligne de commande du shell Unix. L'orateur a commencé par : « vous utilisez toustes le shell? », ce qui évidemment fait rire au Capitole du Libre. Le but du projet est d'améliorer l'invite de la ligne de commande en indiquant davantage d'informations. Le problème est évidemment de ne pas surcharger l'utilisateur. LiquidPrompt, comme ses concurrents, permet d'avoir plusieurs thèmes, et le thème par défaut est dense et plein d'informations. (Y compris la charge de la batterie, ce que je trouve anxiogène.) Pour les démonstrations, c'est spectaculaire. Une autre démonstration montrait un thème où l'invite occupe tout le terminal tellement il y a d'informations.

Un gros avantage de ces systèmes d'invites riches est l'indication de l'état du dépôt Git (ou bien parfois d'autres VCS) où on se trouve. On peut indiquer la branche, bien sûr (combien de fois je me suis planté, me trompant sur la branche où j'étais) mais aussi des suggestions sur les commandes git qui pourraient être pertinentes.

Bon, il y a tellement de systèmes d'invites shell améliorées que c'est un travail en soi d'en choisir une! L'auteur a fait un article de comparaison <<https://linuxfr.org/news/comparaison-critique-de-systemes-d>>

Maïtané Lenoir a parlé de « modération » (terme que je trouve contestable) sur les réseaux sociaux, notamment les « réseaux sociaux capitalistes » (ce qui inclut TikTok, même si ça déplaît au PC chinois). « Je vais vous raconter des trucs traumatiques. » Déjà, elle a noté qu'il fallait bien différencier la politique officiellement affichée, et les moyens qui sont effectivement mis en œuvre pour l'appliquer. (Ce principe doit pouvoir s'appliquer à la politique en général.) Elle a insisté sur l'importance d'une riposte graduée. Si un logiciel ne permet que des réactions simplistes et binaires (supprimer le compte, par exemple), le contrôle sera difficile car on n'aura le choix qu'entre deux mauvaises décisions. Parmi les difficultés (que les réseaux sociaux capitalistes mettent sous le tapis) du contrôle, la langue. Il ne suffit pas de lire la langue utilisée, il faut aussi connaître la culture, pour comprendre allusions, jeux de mots, choix de vocabulaire. Un Français va avoir du mal à évaluer du contenu écrit en français par un Ivoirien ou un Québécois.

Et, bien sûr, le contrôle prend du temps et des efforts et les réseaux sociaux capitalistes, quoi qu'ils prétendent, ne mettent pas assez de moyens. L'IA n'est pas une solution. (Notez que Pharos a la même analyse <<https://www.bortzmeyer.org/pharos-these.html>>.)

L'auteure travaille dans l'équipe de modération de Framasoft, par exemple pour l'instance fédivers Framapiaf <<https://framapiaf.org/>>. Inutile de dire que ce n'est pas facile, cette équipe a dû faire

face à des campagnes s'attaquant à Framasoft et à sa modération (prétendant par exemple qu'il n'y en avait pas). La publication de la charte de modération n'a rien arrangé, les attaques des ultras du contrôle contre Framasoft ont redoublé, tournant au harcèlement. Heureusement qu'en interne, le collectif était soudé.

L'oratrice a beaucoup insisté sur les difficultés pratiques, quotidiennes de la modération : par exemple, beaucoup de signalements ne donnent aucun détail. C'est quoi, le problème, exactement ? Les modérateurs doivent essayer de deviner. Signaleurs, détaillez vos messages ! (Sinon, elle n'en a pas parlé, mais je trouve qu'il y a aussi des signalements malveillants, pour faire taire des ennemis politiques. Les islamistes font souvent cela, par exemple.) Conseil : ne pas oublier le temps et l'énergie consommée. Certes, la modération de Framapiaf n'est pas parfaite (aucune ne l'est) mais un des principes est : « ne pas passer un temps infini sur chaque cas » (surtout que les modérateurs sont bénévoles). Or, les signaleurs sont souvent très exigeants « on veut une action safe tout de suite!!! »

La conclusion : Modération = Pouvoir et Humain = Faillible.

On a eu évidemment une table ronde sur l'IA, où on a surtout parlé de régulation. Celle-ci évidemment ne vise pas forcément qu'à protéger les gens, elle peut avoir des motivations protectionnistes. Florence Sèdes a noté que le récent projet Kyutai <<https://www.numerama.com/tech/1564996-pourquoi-xavier-ni.html>> de Niel et Saadé ne mette en jeu que des sommes « dérisoires ». Toujours dans l'actualité chargée de l'IA, j'ai demandé à ChatGPT les raisons du licenciement spectaculaire de la direction d'OpenAI, la veille du Capitole du Libre, licenciement qui fait parler toute la Silicon Valley mais il m'a répondu qu'il ne savait pas, qu'à sa connaissance, Altman était toujours CEO.

L'autre table ronde portait sur la sobriété, avec notamment Agnès Crepet, qui bosse chez Fairphone. La table ronde a été un peu plus animée car il y avait aussi le député Modem Philippe Latombe, spécialiste des questions numériques, mais qui a voté la loi JO, avec l'extension de la vidéo-surveillance. Indépendamment de la question des droits humains, la vidéo-surveillance est un gouffre énergétique, il s'est fait interpellé là-dessus, a agité le chiffon rouge en disant qu'il ne fallait pas dire vidéo-surveillance mais vidéo-protection (comme on dit dans les comptes rendus de débats « mouvements divers ; protestations »), bref, a contribué à rappeler que la sobriété numérique n'était pas consensuelle. Les JO, qui sont le prétexte pour généraliser cette surveillance, sont une grande source de gaspillage, et Bookynette, qui animait la table ronde, a appelé à boycotter cette fête du fric et du dopage.

Sinon, en matière d'empreinte environnementale du numérique, ne pensez pas qu'à la consommation électrique, note Agnès Crepet. Il y a aussi l'extraction des métaux, dans des conditions désastreuses. Elle a aussi parlé de la réparabilité, sa tante Ginette étant capable de démonter un ordiphone. Ginette est plus habile de ses mains que moi, elle arrive à réparer son Fairphone, alors que j'ai bousillé un Fairphone 2 en essayant d'intervenir à l'intérieur.

Autre participant à la table ronde, Gaël Duval (/e/). Il a parlé de l'obsolescence logicielle. Pour limiter la surconsommation de matériel, il faut aussi maintenir du logiciel qui continue à tourner sur les « vieux » (10 ans) équipements. Par exemple, les vieux trucs 32 bits ont du mal avec des applications qui sont en 64 bits maintenant. Philippe Latombe note que l'État n'est pas forcément cohérent, par exemple en prônant la réutilisation tout en taxant (avec le racket « copie privée ») les téléphones reconditionnés. Agnès Crepet a appelé au « techno-discernement » ; la technique n'est pas forcément utile. Parfois, la solution est de ne pas numériser.

Guillaume a parlé de l'intégration des contributions externes dans le compilateur Rust. Il a expliqué comment les auteurs de Rust géraient ces contributions, entre le désir de favoriser les contributions extérieures (un gros point fort du logiciel libre) et la nécessité de contrôler (pour éviter bogues, voire,

pires, attaques « *supply chain* »). L'équilibre est délicat. Une bonne partie du processus d'examen des contributions est automatisée, via la CI. Si le changement est léger et que la CI passe, ça va plus vite. Sinon, jetez un coup d'œil à la file d'attente de Rust <<https://bors.rust-lang.org/queue/rust>>. Évidemment, il faut que tous les tests passent et, rien que pour les messages d'erreur du compilateur, il y a plusieurs dizaines de milliers de tests. Les machines tournent longtemps ! Si vous voulez contribuer au compilateur Rust, certains tickets sont marqués « pour débutants » et sont a priori plus faciles. Vous pouvez voir pas mal d'articles sur Rust au blog de l'orateur <<https://blog.guillaume-gomez.fr/>>.

Adrien Destugues a fait un exposé très geek, la rétro-ingénierie d'un jeu vidéo où il jouait quand il était petit (le logiciel ludo-éducatif « Lectures Enjeu »), jeu commercial abandonné et qu'il voulait faire revivre. Aucune perspective utilitaire ou financière, c'était juste de l'art pour l'art. Amusant, les chaînes de caractères du jeu sont paramétrées avec un @ qui est remplacé par une marque du féminin si la joueuse a indiqué qu'elle était du genre féminin. "arrivé@" dans le fichier est affiché "arrivé" ou "arrivée" selon le cas.

Un sujet financier ? « La *blockchain* pour le financement du logiciel libre » par Guillaume Poullain, par exemple. Gitcoin est une organisation de financement du logiciel libre (50 M\$ distribués en 2023). Ça utilise le protocole Allo (je ne connais pas). Les gens qui financent ont un droit de vote. L'allocation est quadratique, favorisant le nombre de votes davantage que le montant (pour éviter qu'un millionnaire n'ait autant de poids que mille smicards). 3 715 projets ont été financés, 3,8 M\$ de donations (je n'ai pas vérifié tous les chiffres).

Un des problèmes est évidemment ce sont les projets les plus connus (souvent les plus riches) qui ont le plus de votes. Et, évidemment, certains projets reçoivent peu de votes et donc de fric. Il y a aussi le risque d'attaque Sybil, favorisé par l'allocation quadratique. (Solution : une vérification d'unicité avec un système nommé Passport.)

À noter qu'on ne peut pas retirer l'argent, le flux est dans un seul sens (ce qui évite les bogues style The DAO <<https://www.bortzmeyer.org/the-dao-ethereum-et-une-attaque.html>> qui sont la plaie des fonds d'investissement tournant sur la chaîne de blocs). Bref, mon opinion : ça me semble une bonne utilisation des possibilités de la chaîne de blocs.

Merci à l'école qui prête les locaux gratuitement, aux sponsors. Et surtout aux bénévoles :