



Software Heritage : une archive de 4 milliards de codes sources

C'est une immense bibliothèque qui vient d'ouvrir ses portes virtuelles. Software Heritage vient de lancer son archive en partenariat avec l'Inria, l'Unesco, Github, Microsoft ou encore Google. Y sont référencés et conservés des milliards de fichiers sources.

C'est peut-être une coïncidence, mais elle est presque trop belle pour être vraie. Alors que Microsoft vient d'annoncer le rachat de GitHub, une bibliothèque du code source ouvre ses portes. Au sens figuré bien sûr, puisque l'archive de Software Heritage est en ligne.

Cette initiative soutenue par l'Inria et l'Unesco, qui considère le code source comme patrimoine immatériel de l'Humanité, vient de rendre librement consultables les centaines de milliards de ligne de code qu'elle a collectées depuis sa création.

```
4025     throttled = !list_empty(&cfs_b->throttled_ops_ops);
4026     cfs_b->nr_periods += overrun;
4027
4028     /*
4029      * idle depends on !throttled (for the case of a large backlog, etc)
4030      * we're going inactive then everything else can be done.
4031      */
4032     if (cfs_b->idle && !throttled)
4033         __refill_cfs_bandwidth_runtime(&cfs_b);
4034
4035     if (!throttled) {
4036         /* mark as potentially idle for the upcoming period */
4037         cfs_b->idle = 1;
4038         return 0;
4039     }
4040 }
```

Les logiciels sont notre patrimoine

« Notre objectif à long terme est de collecter tous les logiciels disponibles publiquement sous forme de code source, avec l'historique de leur développement, de les dupliquer massivement pour garantir leur préservation, et de les partager avec tous ceux qui en ont besoin » indique l'archive. La plateforme renferme au moment où nous écrivons ces lignes 4 579 155 739 fichiers sources, un peu plus d'un milliard de commits et 83 millions de projets. Des chiffres qui seront amenés à grandir.

Travail de mémoire sur le logiciel

On y trouve aussi bien le code source de Doom 3 que celui d'Angular.js. Les fichiers sont issus de repositories de GitHub et de feus Gitorious et Google Code, ainsi que des packages Debian et des releases du projet GNU. Ce ne sera pas une surprise : Microsoft, GitHub, Creative Commons, Google ou encore Nokia Bells Lab sont partenaires de Software Heritage. Pour naviguer dans l'archive, un moteur de recherche est disponible ainsi qu'une API.

Inscription gratuite à la newsletter de L'Informaticien.