



Usbek & Rica 16/06/2018

« Nous créons la bibliothèque d'Alexandrie du code source



Phane Montet

#Société #Numérique



« *La bibliothèque d'Alexandrie des codes sources* » : c'est le projet ambitieux de l'UNESCO et de l'INRIA, formé depuis 2015 et dont la première version vient d'être présentée au siège de l'UNESCO le jeudi 7 juin 2018. Cette archive des codes sources nommée Software Heritage a été lancée officiellement le 30 juin 2016 et contient déjà plus de 4 millions de codes sources à l'heure actuelle. Accompagné par de prestigieux sponsors et partenaires comme Microsoft ou la Société Générale, la plateforme d'archivage Software Heritage est encore en développement et estime qu'elle sera stable à partir de 2023, date à laquelle elle aura déjà englouti plus de 17 millions d'euros de budget. Encore à la recherche de sponsors, le directeur du projet Roberto Di Cosmo, chercheur et informaticien, a répondu nos questions concernant l'utilité de Software Heritage, et sur son importance pour les sociétés du futur.

Usbek & Rica : Pouvez-vous nous rappeler ce qu'est un code source ?

Roberto Di Cosmo : On utilise des logiciels tout le temps aujourd'hui, mais on se pose rarement la question de savoir comment ils sont « écrits ». Ces logiciels sont écrits dans des langages de programmation, conçus pour être compréhensibles à la fois par des développeurs, et par les machines. Cela a environ 50 ans, c'est assez récent, et on n'a pas pu s'y habituer comme on l'a fait avec des domaines comme la physique ou la biologie, qu'on connaît depuis des siècles. La plupart des gens ne sont pas encore capables de lire et d'apprécier cette connaissance, même si c'est en train de changer. Il y a des initiatives dans les écoles pour mettre en place des cours





Roberto Di Cosmo © CNRS Le journal

Pourquoi archiver les codes sources ? À qui est destiné cet archivage ?

L'archivage pour les livres, les peintures, les œuvres d'art, tout cela existe déjà, et il existe même une archive de tout l'historique des pages web, mais pas pour les codes sources. C'est étonnant, car les codes sources représentent la clé de la révolution numérique. Je pense qu'on n'y a tout simplement pas pensé avant.

Notre but est de rendre accessibles ces codes sources à toutes les personnes qui savent aujourd'hui lire le code, mais c'est un projet à long terme et le nombre de ces personnes va augmenter au fur et à mesure des années, et cela va devenir une ressource précieuse pour de plus en plus de gens.

« Il existe environ 7 000 langues humaines, vivantes et





Usbek & Rica

Mais il existe énormément de langages de programmation pour écrire des codes sources, et entre temps, ils auront aussi évolué, et peut-être que certains ne seront plus lisibles : comment faire pour les rendre accessibles ?

C'est tout à fait vrai, et je ne sais pas si vous saviez, mais il existe environ 7 000 langues humaines, vivantes et mortes dans le monde, pour 8 000 langages de programmation répertoriés depuis l'invention de l'informatique ! Ils ne sont pas tous complètement différents, mais ils sont nombreux et évoluent très vite, et il y en a certains qu'on n'utilise déjà plus. Mais l'histoire de l'informatique est récente, ce n'est pas comme les hiéroglyphes qui nous étaient complètement inconnus lorsqu'on les a redécouvert. Là, on a de la documentation, des manuels pour les comprendre, et les inventeurs de ces langages sont, pour la plupart, toujours vivants. Prenez par exemple le langage utilisé pour la mission Apollon 11, mise en place par l'équipe dirigée par Margaret Hamilton : on peut toujours lui demander comment elle l'a fait, car elle est toujours vivante !



Margaret Hamilton avec l'intégrale du code source de la mission Apollo 11 © NASA

De plus, dans ces vieux codes sources, les programmeurs avaient l'habitude de mettre énormément de commentaires en langage « naturel » pour expliquer ce qu'ils avaient écrit dans le code source, et cela nous permet de mieux comprendre ce qu'ils ont écrit en langage informatique.

Est-ce que vous sélectionnez les codes sources à archiver ?





assez compacte, et cela ne nous oblige pas à faire une sélection. Et d'un point de vue philosophique, il est important pour nous de stocker tous les logiciels car on ne peut pas savoir à l'avance quel logiciel aura un impact majeur plus tard. Si on n'archive pas tout de suite un logiciel pendant son développement, on ne saura pas qui l'a fait et comment il a été construit, on n'aura pas de traces de ce qui a été fait avant. Donc on préfère tout stocker car s'il y a quelque chose d'intéressant, il sera bien dans notre archive.

« Il existe aujourd'hui plusieurs copies de l'archive dans différents endroits de la planète »

Comment vous assurez-vous que les data centers sur lesquels sont stockés l'archive perdurent sur un temps long ?

On ne prétend pas être les meilleurs, il y aura sûrement des problèmes, c'est naturel. Mais nous avons pris des précautions et il existe aujourd'hui plusieurs copies de l'archive dans différents endroits de la planète. Nous avons trois copies de l'archive : deux sont dans les data centers de l'Inria, et une est sur un cloud de Microsoft dans un autre pays. On travaille aussi sur un réseau de « [miroirs](#) » international, c'est-à-dire qu'il y a d'autres institutions ou d'autres entreprises qui gèrent une copie de l'archive, et qui sont indépendantes de nous. Comme ça, s'il y a un problème avec nos data centers, il y aura toujours une autre copie de l'archive quelque part, intacte et intègre.

Les codes sources évoluent constamment, comment faites-vous pour tenir à jour votre archive ?

On essaye d'être le plus possible à jour, mais ce n'est pas facile quand il y a des dizaines de milliers de projets qui évoluent tous les jours, cela demande de la ressource. C'est pour ça qu'on cherche des sponsors et des partenaires. On travaille avec des plateformes comme [GitHub](#) pour récupérer le plus de codes sources et notamment ceux en développement. L'archive n'est pas seulement la mémoire historique et ancienne des codes sources, c'est aussi un outil magnifique pour observer le développement des logiciels aujourd'hui. On ne veut pas juste regarder les logiciels du passé, on veut surtout améliorer la production des logiciels pour le futur.





```
s_b->idle && !throttled)
goto out_deactivate;

ll_cfs_bandwidth_runtime(cfs_b);

throttled) {
/* mark as potentially idle for the upcoming
cfs_b->idle = 1;
return 0;
```

patrimoine

Nous construisons l'archive universelle des logiciels

Capture d'écran du site Software Heritage

C'est-à-dire ? Comment améliorer les logiciels pour le futur ?

Notre archive est un observatoire absolument unique du développement des logiciels de la planète. Qu'importent les plateformes d'où ils viennent, ces logiciels sont tous présentés de la même manière sur notre archive, cela permet de les regarder tous avec le même regard. Software Heritage est une plateforme où on peut les étudier, trouver des erreurs potentielles, les améliorer, et comprendre comment ils évoluent. Cela va nous permettre de construire des applications et des logiciels meilleurs, et cela représente un intérêt énorme pour la recherche, l'industrie, la science, et on essaye de répondre aux besoins de ces publics très vastes. Des plateformes comme [Google Code](#) ou [Gitorious](#), qui hébergeaient des codes sources, ont fermé (*car elles ne semblaient plus « utiles » à leurs constructeurs, d'autres plateformes comme GitHub les ayant remplacé et gagnant presque tout le trafic*, ndlr). Avec ces plateformes ont disparu tous les codes sources qu'ils contenaient, or notre archive n'est pas destinée à disparaître, les logiciels seront toujours là, avec leurs mises à jour.





Usbek & Rica

découlent, soient reproductibles. Si on peut reproduire une expérience et qu'on arrive au même résultat plusieurs fois, cela peut alors devenir une connaissance scientifique. [Karl Popper](#), philosophe et scientifique, l'a très bien formulé en disant « *les occurrences non-reproductibles n'ont aucun intérêt pour la science* ».

Pour pouvoir refaire ces expériences scientifiques, il faut posséder (...) le logiciel qui a été utilisé pour faire l'expérience

Or aujourd'hui, on utilise de plus en plus de logiciels pour faire des études scientifiques, pour faire des statistiques, ou des expériences que ce soit en physique, en biologie, ou encore en géographie. Et pour pouvoir refaire ces expériences, il faut posséder l'article qui parle des résultats, ainsi que les données et le logiciel qui ont été utilisés pour faire l'expérience. Pour ce qui est des articles, il y en a de plus en plus qui sont accessibles gratuitement, de même pour les données, mais ce n'était pas encore le cas pour les logiciels. Avec Software Heritage, on apporte une réponse à ce besoin et on permet la pérennité de ces logiciels.

Quelle nouvelle mission se donne Software Heritage pour le futur ?

Enormément de travail reste à faire. On est très loin d'avoir fait tout ce qu'on voulait. On ne veut pas juste archiver plus de logiciels, on veut ajouter de nouvelles fonctionnalités intéressantes et utiles pour les utilisateurs. Il y a plus de 4 millions de logiciels répertoriés sur notre site et il est difficile de repérer lesquels peuvent être intéressants. Il nous faut donc trouver une meilleure manière de les répertorier et de les catégoriser, c'est en ce sens que l'on a établi un partenariat avec [Qwant](#), le moteur de recherche. Cela va nous permettre de créer une interface de recherche beaucoup plus sophistiquée, qui permettra de trouver plus facilement ce qu'on veut. Ce sera un moteur de recherche des codes sources.

SUR LE MÊME SUJET :

- > [« On cherche à éveiller la conscience éthique des développeurs et data scientists »](#)
- > [Les archivistes du web en mission sacrée contre Donald Trump](#)
- > [Peut-on éteindre internet ?](#)





Usbek & Rica



Usbek & Rica



Usbek & Rica est le média qui explore le futur. Tous les futurs : ceux qui nous font peur et ceux dont on rêve.



Like

COMMENTAIRES

Commenter...

Commenter

Usbek & Rica

19/06/2017



« Sur son lit de mort, personne ne se dit : "J'aurais aimé passer plus de temps sur Facebook" »

Like Partager

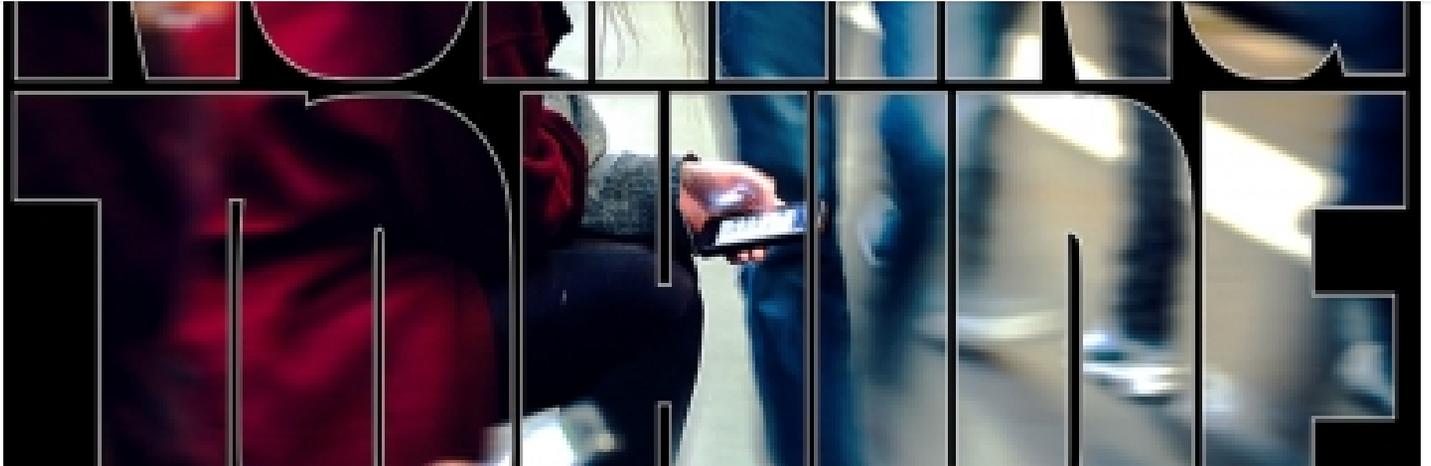
Usbek & Rica

20/09/2017





Usbek & Rica



Pourquoi n'avoir « rien à cacher » n'est pas une raison pour accepter la surveillance de masse

Like Partager

Usbek & Rica

04/06/2017



Vivre sans objets : qui sont les « minimalistes » ?

Like Partager

