



Un entrepôt pour sauver les logiciels



Bibliothèque d'Alexandrie des logiciels », « Wikipédia des codes sources », « Préservation du patrimoine mondial informatique », le président d'Inria, Antoine Petit, ne manquait pas d'emphase, jeudi 30 juin à Paris, pour le lancement d'un projet français de portée mondiale, Software Heritage. Son but est d'archiver la totalité des programmes informatiques de la planète.

« *Tout est logiciel* » , a rappelé Roberto Di Cosmo, professeur d'informatique à l'université Paris-VII, à la tête de ce projet mobilisant moins de dix personnes, financé par Inria. Les programmes sont en effet partout, dans les téléphones, les ordinateurs, les voitures, mais aussi dans la gestion des transports ou de l'énergie.

Mais en matière de logiciels, il faut distinguer deux choses : la partie compréhensible par l'ordinateur, l'exécutable, et la partie lisible par le programmeur, le code source. Souvent cette partie est gardée secrète par le propriétaire et seul l'exécutable est diffusé.

Depuis les années 1980, un mouvement s'est développé : le logiciel libre, dont les licences d'exploitation donnent le droit de lire, d'étudier, de modifier et de diffuser le code source. Software Heritage s'intéresse à cette grande famille qui compte des vedettes comme Firefox, LibreOffice, VLC ou Apache, MySQL, Php (sans qui bon nombre de services Web n'existeraient pas)...

« *Connaissances précieuses* »

Ironie de l'histoire, l'un des opposants historiques aux logiciels libres, Microsoft, est devenu le premier partenaire du projet. « *Je suis assez ému aujourd'hui car, il y a quinze ans, je n'étais pas vraiment l'ami de Microsoft* » , sourit Roberto Di Cosmo, qui, en 1998, avait accusé l'entreprise d'un « hold-up planétaire » dans un livre du même nom.

« *Les logiciels ne sont pas que des outils. Ce sont aussi des connaissances précieuses. Que fait-on pour les protéger ?* » , se demande le chercheur, qui souligne leur « *fragilité* » . En effet, les développeurs utilisent



[Visualiser l'article](#)

pour écrire leurs programmes des sites facilitant le travail à plusieurs et le téléchargement des fameux codes sources.

Cela éparpille la connaissance et surtout ces plates-formes, la plus connue étant GitHub, peuvent disparaître, comme récemment Gitorious ou même Google code. Software Heritage remédie à cela en collectant plusieurs sources. L'intégralité de la collection de GitHub a été récupérée mais aussi, in extremis, celles de Google code et Gitorious. A cela s'ajoutent les fichiers d'un célèbre système d'exploitation, Debian, équivalent de Mac OS X ou de Windows. Au total, 2,6 milliards de fichiers pour 22 millions de programmes seront hébergés grâce à Microsoft et une fondation de l'Académie des sciences des Pays-Bas.

L'intérêt n'est pas seulement la collecte et la préservation de ces « textes », c'est aussi de fournir un outil utile à la recherche (reproduction des résultats, étude des programmes), l'industrie (analyses de sécurité) ou l'éducation (accès à des exemples et références).

Pour l'instant, la plate-forme permet seulement de vérifier si un code source est bien présent, mais pas de chercher ou de naviguer dans ces programmes, ce qui est prévu pour 2017. Les initiateurs comptent sur la communauté pour aider à organiser et à classer cette immense quantité d'informations, afin que l'outil soit vraiment utile.