



Bienvenue dans la bibliothèque d'Alexandrie du logiciel



Il y a un an l'[Inria](#) a imaginé Software Heritage, une bibliothèque d'Alexandrie qui recenserait tous les codes sources des logiciels présent quelque part sur Internet. - D.R.

L'[Institut national de recherche en informatique et en automatique \(Inria\)](#) entend « collecter, organiser, préserver et rendre accessible », le code source de tous les logiciels disponibles sur Internet. Aujourd'hui, tout (ou presque) est logiciel. De votre montre connectée à votre voiture, en passant par les équipements médicaux, les échanges commerciaux et financiers ou encore nos moyens de transports et de communications, les logiciels sont partout et essentiellement sur Internet. Notamment, leur code source.

Ce fichier texte constitué de milliers/millions de lignes de caractères informatiques destinées à être lues par une machine, généralement un ordinateur. Ce sont ces fichiers que l'[Inria](#) entend recenser, référencer et mettre à disposition de l'humanité, de la science et des industriels. Il y a un an l'institut a imaginé Software Heritage, une banque de données mondiale qui recenserait tous les codes sources des logiciels présents quelque part sur Internet.



Les logiciels sont toujours le fruit d'un travail humain, de connaissances humaines, ils appartiennent au patrimoine mondial de l'humanité

Le 30 juin dernier, sûr que le programme pourrait remplir sa mission, l'Inria l'a ouvert au public afin que chacun puisse à la fois retrouver les codes sources référencés mais aussi participer au référencement de nouveaux logiciels.

Une bibliothèque d'Alexandrie du logiciel

« Les logiciels, même s'ils sont destinés à être traités par des machines, à l'origine sont toujours le fruit d'un travail humain, de connaissances humaines, cela appartient donc au patrimoine mondial de l'humanité », explique Stefano Zacchiroli, un des fondateurs et directeur technique de Software Heritage. Pendant un an, lui et toute l'équipe de chercheurs qui s'occupent du projet, ont développé l'interface et vérifié qu'elle était capable de recevoir, sauvegarder et préserver ce savoir afin de le rendre accessible à tous.



L' Institut national de recherche en informatique et en automatique (Inria) entend « collecter, organiser, préserver et rendre accessible », le code source de tous les logiciels disponibles sur Internet. - Inria

« Software Heritage poursuit donc trois objectifs », commente Stefano Zacchiroli. Le premier est de recenser ce patrimoine mondial de l'humanité et d'assurer sa conservation, créant ainsi une sorte de bibliothèque d'Alexandrie du logiciel. « Il faut savoir que si la version exécutable d'un logiciel peut se dégrader dans le temps, le code source, lui, se dégrade beaucoup moins », précise-t-il. Toujours est-il que, pour assurer la sauvegarde des codes sources en cas de problèmes des serveurs ou des disques durs hébergeant le programme, Inria « développe des partenariats » avec des sociétés « spécialisées dans la sauvegarde de données ». Elles hébergeront des programmes « miroirs » de Software Heritage. Des sortes de copies.

Software en quelques chiffres

> **2,5 milliards de « fichiers sources uniques »** ont déjà été archivés dans Software Heritage lors de sa phase « test », pendant un an avant son ouverture au public, le 30 juin dernier. Ce chiffre ne comptabilise évidemment pas les « miroirs », c'est-à-dire les copies de sauvegardes.

> **600 millions de « révisions »** ont été recensées. Il s'agit des modifications apportés au fur et à mesure par de nouveaux développeurs aux codes sources originaux d'un logiciel.



[Visualiser l'article](#)

> **23 millions de « projets de logiciels »** sont inscrits dans Software Heritage. Un logiciel est composé de plusieurs codes sources. L'ensemble de ceux-ci forment un « projet de logiciels » dans lequel généralement plusieurs équipes ont développés différentes parties du logiciel pour l'améliorer.

Organiser « une traçabilité » des codes sources

Le deuxième objectif est de créer « un catalogue des logiciels et de leur code source ». Il s'agit de mettre en place une « traçabilité », afin « d'organiser tout ce savoir contenu dans les logiciels ». On peut alors imaginer qu'un industriel, cherchant un software pour répondre à ses besoins, puisse grâce au « numéro de série du code source », le retrouver dans cette banque de donnée mondiale.

A la différence d'un code ISBN, pour l'identification des livres par exemple, il ne s'agirait pas d'un numéro délivré uniquement par une institution mais par le code source lui-même lorsqu'il est recensé par le programme. De telle sorte que, si l'[Inria](#) venait à ne plus exister, n'importe qui ayant fait une sauvegarde entière de cette banque de données pourrait également générer de nouveaux codes pour d'autres logiciels et ainsi continuer le recensement des fichiers sources.

À lire aussi

Voitures autonomes : piéton ou passager, qui devra mourir ?

La propriété industrielle, accélérateur pour les start-up scientifiques

Enfin, le troisième objectif recherché par les spécialistes de l'[Inria](#), lorsqu'ils ont imaginé Software Heritage, c'est d'aider la recherche et les scientifiques en particulier. Car le logiciel est aussi présent dans toutes les sciences, même les sciences humaines. L'idée est de « capturer les logiciels utilisés par les chercheurs pour leurs expérimentations » et de les référencer en y associant les travaux de recherche, les publications qui y sont associés. Ainsi, lorsqu'un chercheur voudra savoir quel logiciel a utilisé un confrère pour telle ou telle expérience, il pourra le retrouver.

Quid des logiciel malveillants

Reste enfin la question des logiciels malveillants : qu'en faire ? Car Internet c'est aussi le lieu des « chevaux de troie », des « virus » et autres « vers » en tout genre. Autant de sortes de logiciels dont certains noms sont restés célèbres comme « Win32.puper », « Virtual Maid » ou encore « Gator ». « Pas de soucis, répond Stefano Zacchiroli. La partie dangereuse de ces logiciels, c'est la version exécutable. Software Heritage ne recense que les codes sources ».

Et puis, même s'ils sont malveillants, cela reste des créations humaines à l'origine. Comme le reste des logiciels, ils font donc parti de notre patrimoine. Et c'est cet héritage qu'entend laisser Indria, dans la première véritable bibliothèque internationale du logiciel.