

Software Heritage: que faire avec tout le code source du monde?

Roberto Di Cosmo*

*Laboratoire IRIF, université Paris-Diderot, 75205 Paris CEDEX 13
roberto@dicosmo.org,
www.dicosmo.org

Summary

Software Heritage est une initiative à but non lucratif dont l'objectif ambitieux est de collecter, préserver et partager le code source de tous les logiciels jamais écrits, avec leur historique de développement complet, en construisant une base de connaissances logicielle universelle. Software Heritage répond à une variété de besoins : préserver nos connaissances scientifiques et technologiques, améliorer le développement et la réutilisation des logiciels pour la société et l'industrie, favoriser la science ouverte et construire une infrastructure essentielle pour des études logicielles reproductibles à grande échelle. Nous avons déjà collecté plus de 4 milliards de fichiers sources uniques provenant de plus de 80 millions d'origines. Manipuler ce gigantesque ensemble de données est une mission complexe et nécessite de nouvelles approches pour stocker et requêter l'information d'une manière compatible avec la croissance explosive du développement logiciel collaboratif. Dans cette conférence, nous explorons quelques uns des nouveaux défis et opportunités que présente Software Heritage.