

# Software Heritage - Dự án thư viện mã nguồn phần mềm khổng lồ đã ra mắt

[vfossa.vn/vi/news/bao-tuong/Software-Heritage-Du-an-thu-vien-ma-nguon-phan-mem-khong-lo-da-ra-mat-287/](http://vfossa.vn/vi/news/bao-tuong/Software-Heritage-Du-an-thu-vien-ma-nguon-phan-mem-khong-lo-da-ra-mat-287/)

Software Heritage - Dự án thư viện mã nguồn phần mềm khổng lồ đã chính thức ra mắt với tham vọng thu thập, bảo quản, và chia sẻ tất cả các phần mềm có mã nguồn được công bố công khai.



*Software Heritage cho rằng: Phần mềm là chìa khóa thu thập thông tin và là một phần cơ bản của di sản nhân loại*

Dự án Software Heritage (di sản phần mềm) được xây dựng với sứ mệnh thu thập, sắp xếp, bảo tồn và tạo một kênh truy cập tiện lợi nơi các nhà phát triển có thể dễ dàng tìm thấy mã nguồn của tất cả các phần mềm vừa được công bố. Các nhà sáng lập của dự án Software Heritage cho rằng: Phần mềm là chìa khóa thu thập thông tin và là một phần cơ bản và quý giá của di sản nhân loại.

Dự án này được kỳ vọng sẽ tạo được một kênh chia sẻ kiến thức rộng lớn về phần mềm, lập trình cũng như trở thành một nguồn tham khảo chất lượng cho các nhà phát triển trên toàn thế giới. Thậm chí còn hơn thế nữa vì nhờ các công cụ khai phá dữ liệu từ Software Heritage, người ta có thể nhìn thấy quá khứ, hiện tại của tất cả các phần mềm trên thế giới, và biết đâu qua đó sẽ giúp hé mở nhiều thứ cho tương lai của ngành phần mềm.

Dù mới chỉ chính thức khởi động nhưng dự án đã thu lượm được một lượng mã nguồn khổng lồ nhất nhì thế giới từ các kho mã nguồn như Github, Debian, GNU... Hệ thống thống kê số lượng mã nguồn được thu thập của Software Heritage cho biết Software Heritage đã đạt tới 2,6 tỷ file nguồn, gần 600 triệu commit và 22,8 triệu dự án phần mềm.



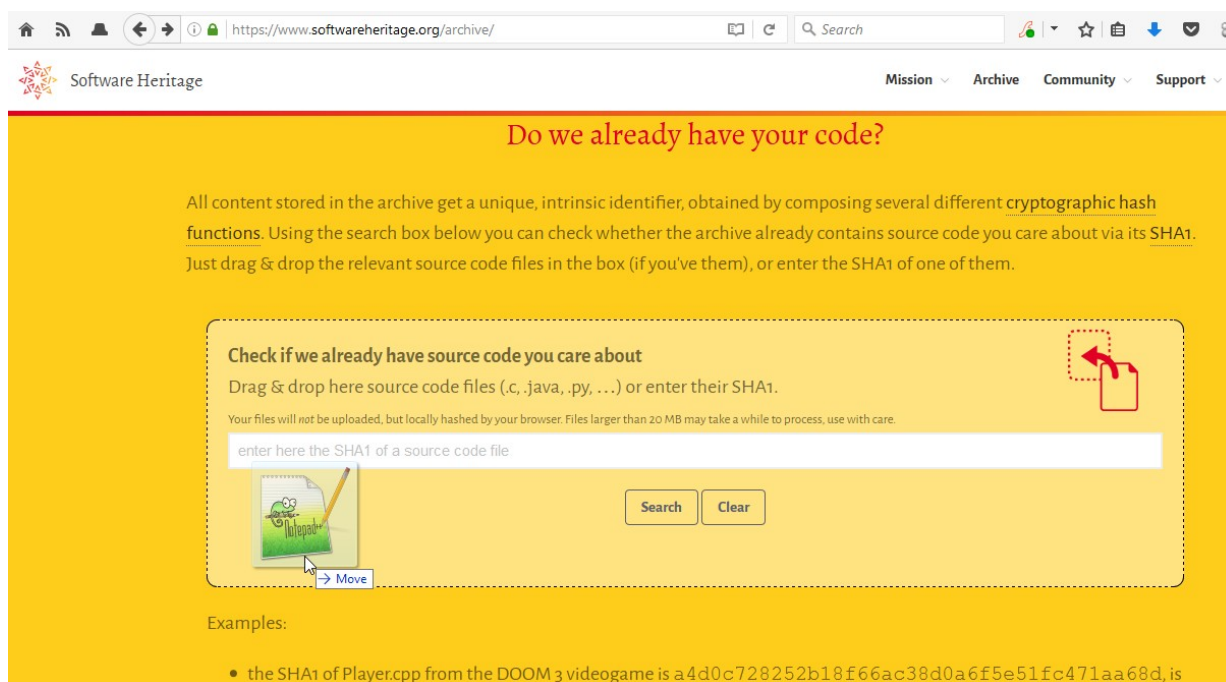
Hình 2: Thông tin thống kê trên Software Heritage

Phần lớn số này bao gồm các phần mềm nguồn mở (không bao gồm các dự án phái sinh) từ GitHub, các gói mã nguồn được cộng đồng phần mềm tự do Debian phân phối (từ 08/2015) cùng một số thu thập được từ dự

án GNU (từ 08/2015).

Dự án được khởi xướng viện nghiên cứu INRIA của Pháp vào năm 2015, đơn vị này cũng chính là nơi cung cấp hầu hết nhân lực cho dự án. Nhà sáng lập kiêm giám đốc Software Heritage là Roberto Di Cosmo. Hiện có 29 tổ chức tuyên bố ủng hộ dự án, đáng chú ý có công ty Microsoft là đối tác cung cấp hạ tầng Azuze và Viện nghiên cứu Dans là đối tác khoa học quốc tế của dự án.

Các nhà phát triển hiện đã có thể đối chiếu xem mã nguồn của họ có mặt ở kho này chưa, tuy nhiên vẫn chưa thể truy cập và sử dụng kho mã nguồn đồ sộ này. Bạn có thể đối chiếu xem mã nguồn của bạn có trên kho chưa bằng cách truy cập địa chỉ [www.softwareheritage.org](http://www.softwareheritage.org), kéo tập tin bất kỳ trong phần mềm của bạn vào khung kiểm tra hoặc nhập mã SHA1 của tập tin (Hình 3).



Hình 3: Kéo các tập tin cần kiểm tra vào khung này để thực hiện việc kiểm tra file.

Hệ thống sẽ trả về kết quả cho biết bao nhiêu % trong số các tập tin này có mặt trên kho của Software Heritage và đó là các tập tin nào (hình 4)

**Check if we already have source code you care about**

Drag & drop here source code files (.c, .java, .py, ...) or enter their SHA1.

Your files will *not* be uploaded, but locally hashed by your browser. Files larger than 20 MB may take a while to process, use with care.

enter here the SHA1 of a source code file

action\_mysql.php, admin.functions.php, admin.menu.php, comment.php, functions.php, global.functions.php, index.html

Search Clear

Found 100.00% of 7 files (click to expand).

FOUND	FILE	SHA1
Yes	action_mysql.php	2308554e9a607611b20defe5c9bf72b8899f5bd8
Yes	admin.functions.php	0ec139e9de6e3ff4545eaa98d40454080494e3ad
Yes	index.html	da39a3ee5e6b4b0d3255bfe95601890afd80709
Yes	global.functions.php	201203108a3ea4d00e9f6519b5875de28332d8f4
Yes	comment.php	22d07d77df9cfe51418de55ac51261a7cc4d00fa
Yes	functions.php	ddd9d867a5fb8f71454eb51bc973eaab9b3261cb
Yes	admin.menu.php	0d05132f0c118b67e21a935d9cf841a487ea8815

Hình 4: Kết quả kiểm tra một số tập tin trong bộ phần mềm nguồn mở [NukeViet](#) của Việt Nam bản mới nhất ([NukeViet CMS 4.0](#)) đã có mặt trên kho code của Software Heritage

Software Heritage đang kêu gọi tài trợ, tham gia vào các cộng đồng nghiên cứu (để giải quyết các thách thức về mặt khoa học của dự án) và code cho Software Heritage (hỗ trợ lập trình cho các phần mềm phục vụ dự án).

Trong lộ trình phát triển, Software Heritage sẽ cho phép các nhà phát triển đưa dự án của mình lên kho của Software Heritage để thực hiện việc chủ động lưu trữ, cho truy xuất nguồn gốc phần mềm/ mã, duyệt các mã nguồn được lưu trữ, tìm kiếm toàn văn mã nguồn và cho phép tải về mã nguồn của các dự án.