

P Software Heritage | Unesco | Università di Bologna

Software patrimonio dell'umanità

Nasce una biblioteca virtuale per preservare il codice sorgente

di **Antonio Dini**

► La matematica è nata nel segreto: Pitagora vietava ai suoi discepoli di divulgare le sue teorie. Al tempo stesso, però, i filosofi greci consideravano le idee universali e, come tali, di proprietà di tutti e liberamente divulgabili. Oltretutto anche il metodo sperimentale della scienza moderna si basa sulla riproducibilità degli esperimenti, cioè sulla sua verificabilità in tutti i laboratori del mondo. Nella tensione tra questi due poli, la segretezza delle idee contrapposta alla loro universalità, cova, a seconda di come la si guardi, la tragedia o il miracolo della scienza moderna. E qui entra in gioco il codice.

Il software infatti, come dice l'inventore del browser grafico e venture capitalist Marc Andreessen, si sta mangiando il mondo. Non solo quello del commercio e degli affari, ma anche quello della scienza sperimentale. Non c'è più settore in cui quel che si fa, si pensa, si scrive, si comunica, si sperimenta, non abbia dietro e dentro il computer. Inclusa la ricerca scientifica.

«Il software – dice Simone Martini, direttore del dipartimento di Informatica dell'Università di Bologna – è alla base di molta parte della scienza moderna. La riproducibilità dei risultati scientifici e l'intero concetto di ricerca scientifica è profondamente legato alla conservazione e allo studio del software sotto forma di codice sorgente». Senza il software, e senza la sua anima – cioè il codice –, in molti casi non è praticamente più possibile ri-

petere determinati esperimenti.

Arriva quindi l'idea di preservare in modo compatto il codice sorgente, ovvero la manifestazione intelligibile e universale del software, che viene poi interpretato oppure compilato per diventare un eseguibile binario. È l'idea di una moderna biblioteca di Alessandria del codice, nata in Francia, ma con un pezzetto di Italia dentro. Si chiama "Software Heritage" (www.softwareheritage.org) e l'ha voluta Inria, l'istituto nazionale francese per l'informatica. Il direttore di Software Heritage è un italiano, Roberto Di Cosmo, mentre il responsabile tecnico è un altro italiano, Stefano Zacchirolì. Dietro Software Heritage ci sono partner tecnologici e sponsor importanti come Microsoft, Intel, Nokia, Huawei. E il dipartimento di Bologna guidato da Martini: «Siamo l'unica università partner di questo progetto».

Bologna ha messo a disposizione del "repository", il deposito del codice di Software Heritage, una serie di computer non più tenuti in prima linea dall'ateneo bolognese, che forniscono invece uno spazio di "staging", un'area di memorizzazione intermedia fra il deposito principale dei dati e gli utenti che vogliono accedervi. Il Disi di Bologna mette a disposizione 65 Terabyte di spazio: con il contributo italiano aumenta la capacità e la fruibilità del progetto senza però che ci sia un innalzamento dei costi di manutenzione, perché i server bolognesi non contengono informazioni critiche. Invece, di aiuto per favorire la fruibilità dei programmi archiviati c'è molto bisogno. Il deposito centrale conserva infatti 3,3 miliardi di documenti sorgente, per un complessivo di più di 58 milioni di progetti software, con quasi 800 milioni di revisioni del codice. Perché il codice sorgente è una materia molto viva, sottoposta a costanti revisioni e versioni, e l'obiettivo è conservare anche le versioni precedenti e successive. La struttura logica dell'archivio è infatti quella

utilizzata dagli sviluppatori quando scrivono codice, e cioè quella dei sistemi di controllo di versione del codice basati su repository distribuiti. È il tipo di ambiente che ogni programmatore, oramai diventato uno sport di squadra e non più un'attività solitaria, frequenta per portare avanti in maniera collaborativa la scrittura del codice.

I 3,3 miliardi di documenti archiviati sono tutti sorgenti, quindi semplici documenti di testo (almeno, la maggior parte). Per essere utilizzati devono essere scaricati e interpretati o compilati ed eseguiti. Ma costituiscono un tesoro unico che lo scorso 3 aprile l'Unesco, in una cerimonia a Parigi a cui ha partecipato anche il presidente François Hollande, ha definito patrimonio universale dell'umanità (è la prima volta che accade per il software).

«L'università – dice Martini – ha tre missioni: ricerca, didattica e trasmissione della cultura, compresa la preservazione del patrimonio. Un obiettivo con cui questa iniziativa si sposa perfettamente». La preservazione del codice sorgente ha un suo significato culturale autonomo. E anche una potenziale valenza scientifica: essendo aperto a tutti può diventare punto di partenza per studi ulteriori. «Possiamo immaginare – prosegue – di fare ricerca partendo da questa base dati e di applicare tecniche prese dallo studio della letteratura tradizionale per cercare genealogie manifeste o nascoste all'interno della scrittura del codice, cercare tic letterari dei programmatori, analoghi concettualmente a quelli degli scrittori nella loro produzione letteraria». Una firma strutturale, cioè iscritta nella struttura linguistica del codice, che permetta, anche all'interno di opere collettive come i software, di recuperare la parte di autorialità del codice, che dopotutto è un manufatto con un altro: un'opera dell'ingegno umano meritevole di essere preservata.

© RIPRODUZIONE RISERVATA

