Piano delle Attività

In recent years we have witnessed an amazing growth of public development platforms which are used to store, develop and share software artefacts.

Even if these platforms are disparate, and there is no common standard for interacting with them, this unprecedented availability of large amounts of source codes and software projects has created new, exciting opportunities and challenges to leverage the advance of Machine Learning and Artificial Intelligence techniques for software development and programming languages practice.

In this context, a significant novelty is the emergence of Software Heritage, an initiative whose goal is to collect, preserve, and share the entire body of software produced in human history in its preferred form of modification---our shared \emph{software commons}.

To this end, Software Heritage is building a revolutionary infrastructure that stores all the history of software development into a unique, giant graph, that abstracts away irrelevant details and provides a uniform data structure, ideal for massive scale analysis.

The ENEA Center in Bologna has recently stepped up to be the first Italian Software Heritage mirror and we recently created in Bologna the Big Code Laboratory with the aim of performing research in big code domain, a domain which is rapidly accelerating and has the potential of being in the near future a disruptive factor for the software industry.

The Lab involves persons from ENEA, INRIA and Software Heritage, University of Bologna.

In this context, the candidate will identify one specific big code research theme of interest and will develop it, using as starting dataset the Software Heritage archive. Example of themes are: formalization of software licenses and their automatic recognition; ML techniques for predicting the presence or errors in source code; use of ML for predicting the author of a source code snippet.