

Efficient program representation for code comprehension

Tibor Brunner, Zoltán Porkoláb

Department of Programming Languages and Compilers, Eötvös Loránd University

`bruntib@ik.elte.hu`, `gsd@elte.hu`

In industrial environments where a legacy code base consists of multiple million lines of code, it is a challenging task to get a proper overview of the software. Code comprehension tools aim to support this process by providing fast navigation opportunities and visualizations like inheritance diagrams, function call chain detection, pointer analysis, etc. CodeCompass is a code comprehension framework that gathers statically collected information through extensible plug-ins like language parsers, metrics measurements or version control systems.

The challenging task of such a system is to store this huge amount of data in a database as concise as possible. It is a primary target to get fast access to this data in order to provide an efficient service for an extensive overview of the whole legacy system. In this paper we describe the data model that CodeCompass uses to store this wide range of information.