

Abstract

We present a supervised clustering environment for the analysis and restructuring of the legacy software systems into cohesive components. The proposed technique is novel in the sense that it pre-processes the relations among the system entities using data mining techniques in order to extract groups of maximally associated system entities. This property can be further used to define similarity metrics between two individual entities or between two groups of entities. Consequently, a two-phase supervised clustering process is used to incrementally generate new clusters. The proposed architectural recovery process is hierarchical in the sense that the software system is first decomposed into subsystems of files and then each subsystem can be further decomposed into modules of system entities such as functions, datatypes, and global variables. An interactive environment provides evaluation metrics to assess the quality of the obtained clusters according to their cohesion and coupling properties. Experimental results obtained by applying the proposed clustering technique on six open source systems are presented.

⁰Portions reprinted, with permission, from the paper “Component Clustering Based on Maximal Association” in conference IEEE WCRE-2001, October 2001. (© 2001 IEEE).

For this work we collaborate with IBM Canada Ltd. Laboratory - Center for Advanced Studies (Toronto) and the National Research Council of Canada.