EXTENDED ABSTRACT

Based on the ISO/IEC 9126 standard for software product quality, the Software Improvement Group (SIG) developed a tool to automatically analyze software maintainability, the Software Analysis Toolkit (SAT). The SAT is used in the SIG consultancy services to support management decisions. SIG has experience analyzing software systems in the financial, insurance, government, logistical and IT domain. We will report on the usage of SAT to analyze the maintainability of two simulators used in the space domain which have been developed with rigorous software processes: EuroSim and SimSat.

The SAT measures maintainability using a five-stars rating: five stars are used for very good quality and one star for very-low quality. The maintainability model is layered and can be decomposed into sub-characteristics, as defined in the ISO/IEC 9126 standard, and then into product properties defined by SIG [1]. The product properties are volume, duplication, unit complexity, unit size, unit interfacing, and testing. These product properties are automatically measured from the source code. For each product property a rating is derived by comparing measurements to thresholds, which have been calibrated with a large number of systems [2,3,4]. The calibration process ensures that five stars represents 5% of all systems (the best quality), four, three and two stars represent each 30% of the systems, and one star represents the rest 5% of the systems (the worst quality).

To use the SAT custom configurations were developed. These configurations specify individual projects needs, such as the programming languages to analyze and how to distinguish production and test code. Custom filters were also defined to ignore generated code, code examples and library code. For each system, the SAT was executed generating a database with all the measurements and deploying a web application. The web application displays the key indicators for software quality in a dashboard and allows drilling down to measurements at the lowest level.

The SAT maintainability analysis revealed a rating of three stars for EuroSim and two stars for SimSat. In-depth analysis of both systems revealed problems in most of the product properties. As example, for duplication the SAT measures the number of identical lines of code and identifies clones (contiguous blocks of identical code). Besides showing the existence of the clones, the SAT identifies the files in which the clones are found, and the start and end lines of each clone. Substantial duplication implies high maintenance costs making bug fixing and testing harder. For EuroSim 7.1% of the overall code is duplicated, while for SimSat this value amounts to 10.4%. For both systems we detected several identical files and several clones larger than 100 lines of code. Curiously, we discovered that for both systems several clones found were due to the (different) implementations of the ESA Simulation Model Portability library (SMP). As an additional example, for unit complexity the SAT measures the number of decisions per function (McCabe metric). Functions with higher complexity are harder to understand and, consequently, harder to maintain. For EuroSim, 42% of the overall code contains moderate and (very) high complexity, while for SimSat this values amounts to 20%. However, while in EuroSim the complexity is spread throughout the system in SimSat it is localized in just a dozen of files.

Additionally, the SAT can be used to support management decisions through the identification and quantification of technical problems. In EuroSim, the complexity analysis helped to uncover that specific teams were responsible for the modules with lower quality. In SimSat, the unusual duplication, size and complexity found for recently developed code, helped to uncover that the team involved with that part of the project lack familiarization with the technology used. In both cases, a decision could be made not to use these teams, or to provide the teams with specific training for producing higher quality code. A more detailed analysis of EuroSim and SimSat quality, and lessons learned can be found in [5].
To conclude, the SAT provides a pragmatic and practically feasible approach to measure product quality, which can be used to support management decisions. The quality model implemented in SAT uses a small set of key metrics whose measurements are aggregated in a five-star ranking. By design, the metrics are operational, understandable and generic. The star rating enables comparison and easy communication between all the stakeholders of a project. The analyses of EuroSim and SimSat revealed that although both systems were developed with rigorous software processes, implementing strict standards, quality problems were found. Hence, it is important to have mechanisms to identify and quantify quality problems during both development and maintenance activities.

REFERENCES