

DISS. ETH No. 21166

Automatic Translation and Object-Oriented Reengineering of Legacy Code

A dissertation submitted to
ETH ZÜRICH

for the degree of
Doctor of Sciences

presented by
MARCO TRUDEL

Master of Science ETH in Informatik, ETH Zürich, Switzerland
Ingenieur FH in Informationstechnologie, ZHW, Switzerland

born
July 18th, 1983

citizen of
Männedorf, ZH

accepted on the recommendation of

Prof. Dr. Bertrand Meyer, examiner
Prof. Dr. Hausi A. Müller, co-examiner
Prof. Dr. Richard Paige, co-examiner
Dr. Manuel Oriol, co-examiner

2013

ABSTRACT

Can we reuse some of the huge code-base developed in C to take advantage of modern programming language features such as type safety, object-orientation, and contracts? This dissertation presents a source-to-source translation and object-oriented reengineering of C code into Eiffel, a modern object-oriented programming language, and the supporting tool C2Eif. The migration is completely automatic and supports the entire C language (ANSI, as well as many GNU C Compiler extensions) as used in practice, including pointer arithmetic, usage of native system libraries and inlined assembly code. Eiffel programs are created exhibiting elements of good object-oriented design, such as low coupling and high cohesion of classes, and proper encapsulation. The programs also leverage advanced features such as inheritance, contracts, and exceptions to achieve a better usability and a clearer design. Our experiments show that C2Eif can handle C applications and libraries of significant size (such as `vim` and `libgs1`), as well as challenging benchmarks such as the GCC torture tests. The produced Eiffel code is functionally equivalent to the original C code, and takes advantage of some of Eiffel's features to produce safe and easy-to-debug programs.

We have also investigated the related problem of automatically translating source code between different modern environments, in particular from Java to Eiffel. The translation has been formalized in order to increase confidence in its correctness, and the usability of the supporting tool J2Eif has been evaluated on four programs of varying complexity.

ZUSAMMENFASSUNG

Ist es möglich einen Teil der riesigen in C entwickelten Code-Basis wiederzuverwenden um die Vorteile von modernen Programmiersprachen wie Typsicherheit, Objektorientierung und Verträge nutzen zu können? Diese Dissertation präsentiert eine Quelltext-zu-Quelltext Übersetzung und objektorientierte Umstrukturierung von C Code nach Eiffel, eine moderne objektorientierte Programmiersprache, und das unterstützende Tool C2Eif. Die Migration ist völlig automatisch und unterstützt die gesamte C Sprache (ANSI sowie viele GNU C Compiler Erweiterungen) wie sie in der Praxis verwendet wird, einschliesslich Zeigerarithmetik, der Benutzung von nativen System-Bibliotheken und Inline-Assembler-Code. Eiffel Programme werden erstellt die Elemente von gutem objektorientiertem Design besitzen wie lose Kopp lung und starke Bindung von Klassen und geeignete Kapselung. Die Pro grammme setzen auch fortschrittliche Merkmale ein wie Vererbung, Verträge und Ausnahmen um eine bessere Benutzerfreundlichkeit und ein klareres Design zu erreichen. Unsere Experimente zeigen dass C2Eif C Anwendungen und Bibliotheken von signifikanter Grösse (wie z.B. `vim` und `libgs1`) sowie anspruchsvolle Benchmarks wie die GCC Folter Tests bearbeiten kann. Der erzeugte Eiffel Quelltext ist funktional äquivalent zu dem ursprünglichen C Quelltext und nutzt einige Eigenschaften von Eiffel um sichere und einfach zu korrigierende Programme zu erstellen.

Wir haben ebenfalls das ähnliche Problem einer automatischen Quelltext-zu-Quelltext Übersetzung zwischen verschiedenen modernen Sprachen untersucht, namentlich von Java zu Eiffel. Die Übersetzung wurde formalisiert um das Vertrauen in die Korrektheit zu erhöhen und die Benutzbarkeit des unterstützenden Tools J2Eif wurde an vier Programmen mit unterschiedlicher Komplexität getestet.