# DWARF2 デバッグ情報を用いた プログラム理解ツールの設計と実現

210051 鈴木 朝也

2004 年 2 月 16 日

C 言語は，安全性などに問題があるが，現在でも幅広く使われている．その理由は，アセンブリ言語による記述が必要な場面でも，C 言語を使えば，実行速度や消費メモリをあまり犠牲にせずに，開発効率や移植性を大幅に向上できるからである．

C 言語は，OS や言語処理系など，コンピュータシステムの基盤となるソフトウェアに使われることが多いため，C で記述されたコードの信頼性向上は重要な課題である．しかし，C 言語の性質上，たとえばアドレス演算を使ってハードウェアの固定アドレスへのアクセスを許すために，C 言語処理系のコンパイル時・実行時のチェックは甘い．このため，C のコードのバグを減らし，より安全にするためには CASE ツールの提案が必須である．

柔軟性の高い CASE ツールをより効率よく開発するには，CASE ツールがデータを共通に使えるように共通フォーマットを定義すること，つまりデータ統合が有効であるとされ，すでに PCTE や CDIF などが研究・開発されてきた．しかし，これらは未だ，仕様の複雑さなどのため広く使われていない．

近年，共通データフォーマットとして XML が注目されている．XML によるソースコードレベルのデータ統合の研究には Sapid，ACML，JavaML，GCC-XML などがある．これらの技術の長所は，オープンで柔軟性の高い CASE ツールを効率よく開発できることである．しかし，欠点もある．この方式の大きな欠点は，(1) ライブラリなどソースコードがないソフトウェアや (2) C 言語処理系の差異，への対応が難しいことである．後者に関しては，例えば，GCC は独自拡張として asm 構文や __attribute__ 構文を持っているが，Sapid と ACML では字句・構文解析時にこれらをスキップして対処しており，すべての独自拡張に対応しているわけではない．この結果，GCC （や他のコンパイラ）ではコンパイル可能だが，Sapid や ACML では，意味をなさない，あるいは処理できないプログラムが数多く存在する．また，GCC のバージョンごとにスキップ処理を調整する必要もあり，手間がかかる．

ソースコードレベルのデータ統合とは，ソースコードの静的な情報，特に構文情報，型情報，シンボル情報の共通フォーマットを与えることである．共通フォーマットを持つことは，例えば XML を用いた研究では，開発効率などの観点からその有効性が確認されつつある．

しかし，ソースコードレベルデータ統合では，ソースコードを解析するため，各ツールが解析器を用意しなければならない．このため，C のソースコードの構文解析器と意味解析器は，コンパイラの独自拡張機能，新しい C99 規格，未規定動作と処理系定義動作に対応する必要がある．これに全て対応するにはコストがかかるし，対応してないツールも存在する．

その問題を解決する一つの手段として，バイナリコードレベルのデータ統合がある．バイナリレベルで動作するツールは多い．例えば，Purify や PureCoverage はソースコードではなく，バイナリコードを検証･修正するツールである．バイナリレベル統合は，このようなバイナリレベルのツー

ルに共通フォーマットを提供する．バイナリコードレベルデータ統合は，適用範囲がひろいなど，多くの利点がある．

そこで，本研究ではバイナリレベルのデータ統合の有用性およびバイナリレベル情報の応用可能性について調べる．

バイナリレベルの情報の一つにデバッグ情報がある．デバッグ情報とは，バイナリ中に含まれている，デバッガに必要な情報のことである．例えば，局所変数・ユーザ定義型・行番号・スコープ・スタックフレームなどがある．これらの情報はデバッガ以外の CASE ツールにとっても有用であると我々は考える．

また，デバッグ情報のフォーマットの一つに DWARF2 があるが，これは GCC・GDB などがサポートしているので，これを利用した CASE ツールは適用範囲が広いことが期待できる．

しかし，バイナリ情報はエンコードしなければ使えず，また DWARF2 デバッグ情報を readelf コマンドで取り出した結果は，可読性はあるが，機械で処理するのには向いていない．そこで，本研究では，バイナリレベル統合のプラットフォームの 1 つとして開発された DWARF2-XML を利用した．DWARF2-XML は readelf の DWARF2 デバッグ情報の出力に解釈を加え構造化した XML 文書である．XML は加工容易性がある．そのため，XML を用いれば，開発コストが低くなる可能性がある．

本研究は，DWARF2 デバッグ情報を XML ベースで提供する DWARF2XML 文書を加工し，相互参照データを作成するクロスリファレンサの設計および実装を行う．ここでの相互参照データとは HTML のハイパーリンクを利用して，プログラムの特定の要素を参照できるようにしたものである．そして，既存のソースコードレベルのクロスリファレンサと実行速度や機能の面で比較する予備実験を行う．

これにより，XML によるバイナリレベルのデータ統合の有用性，バイナリ情報の応用可能性を確かめる．

また本研究では，デバッグ情報を利用したクロスリファレンサを既存のソースコードレベルのクロスリファレンサである GNU GLOBAL を融合させたハイブリッドクロスリファレンサの設計および実装を行い，バイナリレベルの情報の他 CASE ツールへの利用可能性をしらべた．

さらに，このクロスリファレンサ (rxref) を用いて予備実験を行った．

その結果，この予備実験に限って言えば，柔軟で適用範囲の広いクロスリファレンサを短時間（一人月）で開発できた．残念ながら，XML と Ruby を用いたため実行速度は既存のものより遅かった．