

B-048

Java をターゲットにした自国語プログラミングにおける自動変換方式の改良
Improvements to automatic conversion of native language programs from/to Java programs滝本 修平†
Shuhei Takimoto野口健一郎‡
Kenichiro Noguchi

1. はじめに

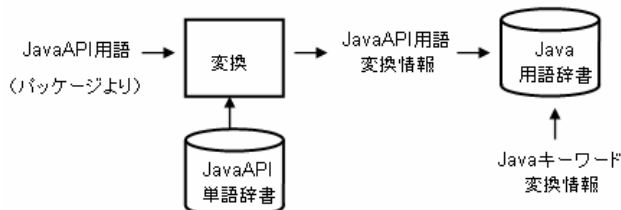
これまで、プログラムのキーワード類を自国語表記から正規の Java 表記に変換することにより、Java プログラムの全てを自国語で記述できるようにする実験を行い[1]、さらに多数の Java API 用語についても、自国語表記（日本語表記）から Java 表記へと変換する実験[2]を行った。本実験では、さらに JavaAPI 用語とユーザ定義用語を区別して変換できる仕掛けの実現などの改良を行い、自国語プログラミングをより実用性のあるものにした。なお今回の実験では、自国語は日本語のみを対象にした。

2. システムの概要

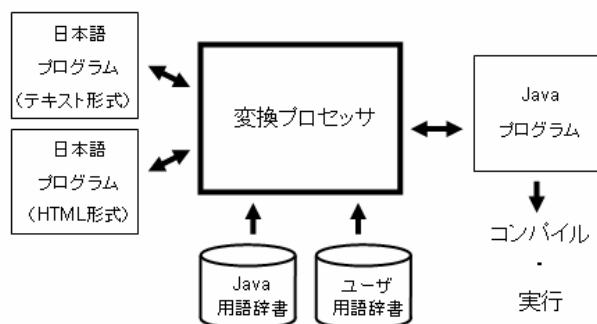
システムの構成図を図 1 に示す。

●JavaAPI 用語の変換情報は、あらかじめ手で作成した JavaAPI 単語辞書を用いて自動生成する。その結果を Java 用語辞書へ登録する。

●自国語（日本語）プログラムから Java プログラムへの変換、および逆に Java プログラムから自国語プログラムへの変換は、変換プロセッサにより自動的に行う。その際、Java キーワードと JavaAPI 用語の変換情報を保持した Java 用語辞書と、ユーザ定義用語の変換情報を保持したユーザ用語辞書を使う。



(a) Java 用語辞書の作成

(b) プログラムの変換
図 1 システムの構成図

3. 研究課題

- ① JavaAPI 用語とユーザ定義用語の区別
- ② JavaAPI 用語変換情報の JavaAPI 単語辞書からの自動生成（自然な日本語用語への変換、変換の一意性の確保）
- ③ 自国語プログラムの見やすい表示形式

4. 実現方式

4. 1 変換用辞書の作成

(1) JavaAPI 単語辞書の準備

JavaAPI 用語の数は膨大であり、JavaAPI 用語の変換情報を手作りするのは、大変な労力を要するだけでなく、首尾一貫した変換情報の作成は難しい、という問題点がある。そこで、JavaAPI 用語の多くは複数の単語の組合せであることに着目し、JavaAPI 用語で使われている単語を抜き出して、それに対応する日本語表現を与えた JavaAPI 単語辞書を作成した。この辞書には品詞情報も持たせた。辞書形式はデータと文書の 2 面性を持つ XML を採用した。例を図 2 に示す。

```
<word>
  <english>set</english>
  <mean type="名詞">セット</mean>
  <mean type="動詞">設定する</mean>
  <mean type="過去分詞">設定された</mean>
</word>
```

図 2 JavaAPI 単語辞書の一部

(2) JavaAPI 用語変換情報の生成

まず Java のパッケージから JavaAPI 用語を抽出し、各用語について JavaAPI 単語辞書を用いて対応する日本語用語を自動生成した。変換アルゴリズムを図 3 に示す。生成する日本語用語は、品詞の前後関係によって決定する。変換例を示す。

例) `getDeclaredClass` 宣言されたクラスを得る
`getDeclaringClass` 宣言しているクラスを得る
`classDeclaring` 宣言しているクラス

単語の品詞情報では適当な変換ができない用語、つまり品詞が文脈依存で決まる用語などについては、例外として手作業で登録した。

例) `readOnly` 読み込み専用（名詞）
`fillIn` 埋め込む（動詞）
`IgnoreCase` 大文字小文字を無視して（副詞）

(3) ユーザ用語辞書への用語登録

ユーザ定義用語の変換情報の登録は利用者が行う。

† 神奈川大学理学部情報科学科（現在 日立公共システムエンジニアリング（株））

‡ 神奈川大学理学部情報科学科

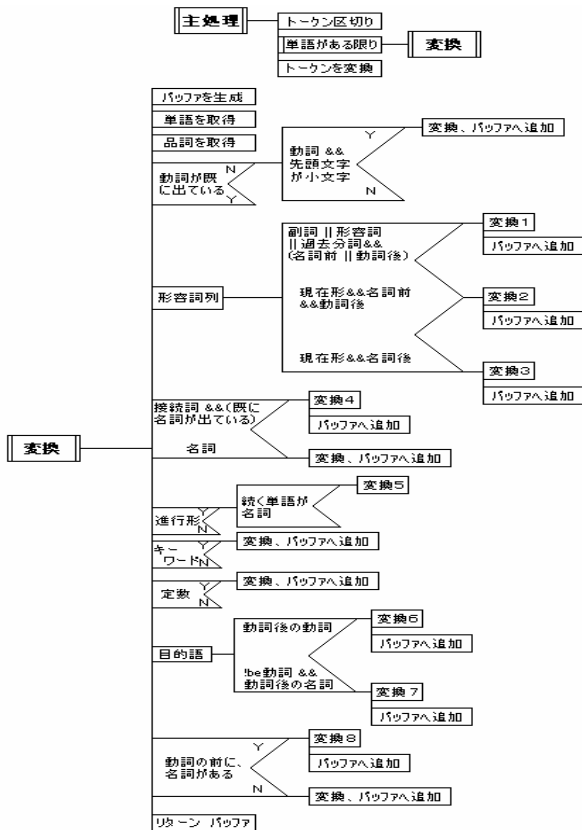


図3 単語辞書を用いた用語変換アルゴリズム

```
import java.util.Collections;
import java.util.ArrayList;
import java.util.List;

公開 クラス 単語 実装 比較可能 {
    非公開 文字列 英語 = "";
    非公開 配列リスト 意味リスト = 新規の 配列リスト();

    公開 単語() {}

    公開 空の 英語を設定する(文字列 英語) {
        自分.英語 = 英語.小文字へ();
    }

    公開 文字列 英語を得る() {
        戻す 英語;
    }

    公開 プーリアン 意味を追加する(意味 いみ) {
        戻す 意味リスト.追加する(いみ);
    }
}
```

```
import java.util.Collections;
import java.util.ArrayList;
import java.util.List;

public class Word implements Comparable {
    private String english = "";
    private ArrayList meanList = new ArrayList();

    public Word() {}

    public void setEnglish(String english) {
        this.english = english.toLowerCase();
    }

    public String getEnglish() {
        return english;
    }

    public boolean addMean(Mean mean) {
        return meanList.add(mean);
    }
}
```

図4 変換したプログラムの一部

1851 行のプログラムにつき変換を行い、全て変換できた。例を図4に示す。

4.2 本国語プログラム/Java プログラムの変換

(1) 変換の概要

変換プロセッサで、プログラム中の用語(トークン)を切り出して、2種類の辞書を参照して変換する。

(2) import 文による JavaAPI 用語の区別

用語変換において、キーワード以外の用語が JavaAPI 用語かユーザ定義用語かの区別が必要になる。そのために、import 文よりクラス名を得て、そのクラスの全 API 用語を取得し、プログラムで使用している用語が JavaAPI 用語に該当するかどうかを判別した。なお、このために、日本語プログラムにおいても import 文だけは英語のまま書くようにした。

4.3 見やすい表示形式

本国語プログラムの出力形式として、テキスト形式のほかに HTML 形式もサポートした。これにより日本語プログラムのキーワード、API 用語、利用者定義用語を、文字のフォント種別や色で区別でき、プログラムの見やすさを実現した。

5. 結果

今回の実験では Java の主要な 7 パッケージを対象に、JavaAPI 単語辞書へ 1250 個の単語を登録した。この結果 JavaAPI 用語 1827 語の変換情報を生成できた。なお、自動生成できなかった用語 38 個は手作業で登録した。

実験で実装した Java プログラム、20 ファイル合計

6. 考察

(1) 以前の変換プロセッサよりも、Java 標準 API の変換範囲が拡大した。

(2) import 文から JavaAPI 用語を区別することで、ユーザ定義の用語も変換してしまうという副作用がなくなった。

(3) プログラムを HTML 形式で出力して表示することで、プログラムの読みやすさが増す。

7. 今後の課題

(1) 用語自動変換の改良 : in や to などの前置詞の変換など、改良の余地がある。

(2) 対象用語の範囲拡大 : 今回は定数について対応できていない。また、対象パッケージの拡大も必要。

参考文献

[1] 鈴木康彦、野口健一郎、後藤英一 : Java 言語をターゲットにした本国語プログラミングの実験、平成 11 年度情報処理学会 第 58 回全国大会
 [2] 松下尊信、野口健一郎 : Java をターゲットにした本国語プログラミングにおける辞書を用いた自動変換、FIT (情報科学技術フォーラム) 2002