

利用実績に基づくソフトウェア部品検索システム SPARS-J

横森 励士[†]

山本 哲男^{††}

松下 誠^{†††}

楠本 真二^{†††}

井上 克郎^{††}

あらまし 本論文では、利用実績に基づいたソフトウェア部品の解析・検索システム SPARS-J(Software Product Archive, analysis and Retrieval System for Java) の紹介を行う。SPARS-J は Java を対象とした部品検索システムで、収集されたソフトウェア部品の集合に対して利用関係に基づいて各部品の重要度を計算し、検索結果の順位付けを行う。これにより、実際によく利用されている部品が検索結果の上位に現れるため、検索キーと関連したソフトウェアのソースコード等を容易に検索できる。

キーワード ソフトウェア部品, 部品検索システム, 再利用, 利用関係

1. ま え が き

ソフトウェアの大規模化と複雑化に伴い、高品質なソフトウェアを一定期間内に効率良く開発することが重要になってきている。これを実現するため、近年のソフトウェア開発において再利用を用いた開発がよく行われている。再利用とは、既存のソフトウェア部品を同一システム内や他のシステムで利用することを指し、開発期間の短縮や品質向上を期待できるといわれている。ソフトウェアの再利用による効果を最大限に引き出すためには、開発者が今から開発しようとするソフトウェアに必要な部品およびライブラリに関する知識を得ることが重要になる。しかし、知識の共有が満足になされていないために、同種のソフトウェアが別々の場所で独立して開発されていることも多い。

一方でインターネットの普及により、SourceForge [1] などのソフトウェア開発に関する情報を交換するコミュニティが誕生し、大量のソフトウェアソースコードが容易に入手できるようになった。これらの公開されている大量の部品の中から、開発者の必要としている機能を持つ部品、その機能の使い方を示している部品のような、再利用に有益な情報を提供する検索システムを実現することで、知識の共有が実現でき、再利用を促進することができると思われる。

そこで本論文では、我々が提案する利用実績に基づくソフトウェア部品検索システム *SPARS-J*(Software

Product Archiving, analyzing, and Retrieving System for Java) について紹介する。SPARS-J は収集してきた Java のプログラムを解析し、リポジトリに保存し、利用実績に基づいた評価値 (Component Rank [2]) により順位付けを行う。開発者は必要な部品に関する情報を検索キーとして入力することで、検索にヒットした部品に関する情報を得ることができる。結果の出力における表示順位として Component Rank を利用することで、開発者の必要としている機能を持つ部品の中でも利用実績の高い汎用的な部品を検索結果の上位に置くことができる。

2. SPARS-J

SPARS-J は Java プログラムを対象とした部品検索システムで、収集された部品に対して解析およびインデックス付けを行うことで検索機能を提供する。SPARS-J は現在のところキーワード検索に対応しており、検索結果はクラス単位で表示される。検索結果を表示する際、検索された部品を評価し順位付けし選別して表示する仕組みが必要となるが、SPARS-J においては、部品間の利用関係を元に定められた Component Rank [2] に基づいて順位付けを行うことで、利用実績の高い汎用的な部品の取得を容易にしている。図 1 に SPARS-J の構成を示す。SPARS-J は、入力されたソースコードに対してリポジトリを構築するリポジトリ構築部と、検索に必要な情報を保存するリポジトリ、リポジトリから情報を取得することで部品の検索を行う部品検索部から構成されている。以下、リポジトリ構築部および部品検索部について説明する。

2.1 リポジトリ構築部

リポジトリ構築部では、検索時に必要なデータを提供するために、次の手順で入力された Java ソースコードからリポジトリの構築を行う。部品のグループ化、Component Rank の計算については後述する。

[†] 大阪大学大学院基礎工学研究科, 大阪府 Graduate School of Engineering Science, Osaka University, Toyonaka, Osaka, 560-8531 Japan

^{††} 科学技術振興事業団, 埼玉県 Japan Science and Technology Corporation, 4-1-8, Honmachi, Kawaguchi, Saitama, 332-8531 Japan

^{†††} 大阪大学大学院情報科学研究科, 大阪府 Graduate School of Information Science and Technology, Osaka University, Toyonaka, 560-8531 Japan

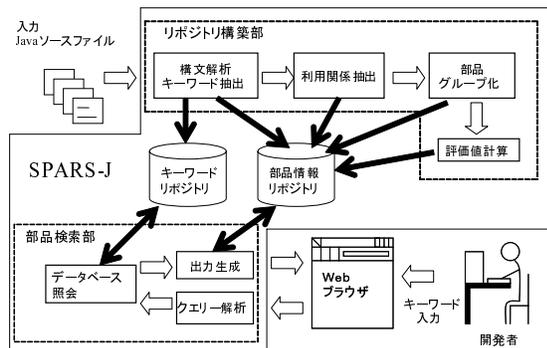


図1 SPARS-Jの構成

(1) 入力されたソースコードに対して構文解析を行い、利用クラス名や、LOC、トークン数などのメトリクスを抽出し部品情報リポジトリに保存する。

(2) ソースコード中で出現する単語に対してインデックス付けを行い、キーワードリポジトリを構築する。

(3) 各部品が利用しているクラスが実際にどのクラスかを意味解析により判定し、利用関係情報として部品情報リポジトリに保存する。SPARS-Jでは、クラス継承、メソッド呼び出し、フィールド参照、抽象クラスの実装を利用関係とみなしている。

(4) 似た部品を一つの部品として扱うために、メトリクスを用いて部品のグループ化を行う。

(5) 利用関係から各部品の Component Rank を求め、部品情報リポジトリに保存する。

2.1.1 部品のグループ化

一般に部品の集合には、コピーした部品やコピーして一部を変更しただけの部品が数多く存在する。異なるシステムをまたいで全く同じもしくはほとんど似た部品が現れる場合、それらの部品が再利用されたのではないかと推測することができる。そこで、取得したメトリクスから類似部品を判定し、それらをまとめて一つの部品群とみなすことでグループ化を行う。その際、それぞれの部品への辺が一つの部品群への辺とみなされるため、コピーされた部品への評価を高くすることができる。

2.1.2 Component Rank の計算

SPARS-Jでは、部品間の利用関係を元に評価値 (Component Rank) を計算する。計算の際には、各部品を頂点、部品間の利用関係を利用する側から利用される側への有向辺として、部品間の関係を部品グラフとして表現する。この時、開発者はある部品を参照した後にグラフ中の有向辺に沿って利用関係のある部品の一つを参照すると仮定することで、我々はこの部品グラフを開発者の閲覧行動に関するマルコフ連鎖モデルとみなしている。このモデルにおける定常状態に

おいて各部品が参照されている確率を求め、その値を Component Rank としている。Component Rank を用いることで、ただ単に利用数が多い部品だけでなく利用数が多い部品が利用している部品も重要であると評価することができる。

2.2 部品検索部

部品検索部では、部品情報リポジトリおよびキーワードリポジトリを用いて部品の検索を行う。現在のシステムでは、基本検索として全てのソースコードからのキーワード検索を実現している。さらに検索条件を指定することで、コメントのみにキーワードが現れる部品を省く、クラスやメソッドの定義にキーワードが出現する部品のみを表示する、などの詳細な検索を行うことができる。検索の手順は以下のようになっている。

(1) 開発者はブラウザを通じてキーワードを入力することで、部品検索部にクエリーを投げる。

(2) 部品検索部はクエリーをキーワードの集合に分解し、キーワードリポジトリに対してそれぞれのキーワードが出現する部品を照会する。

(3) 部品検索部は検索条件およびクエリーにマッチした部品を Component Rank の順に並べ、それを検索結果として出力する。

(4) 開発者はブラウザを通じて検索結果を受け取る。開発者はさらにキーワードを追加することでより詳しい検索を行うことや、検索結果として表示された部品に関して、ソースコードやその部品のメトリクスなどの部品の詳しい情報を取得することができる。

3. まとめ

本論文では、利用実績に基づくソフトウェア部品検索システム SPARS-J を紹介した。SPARS-J では、検索機能を提供する際に部品間の利用関係を元に定められた Component Rank に基づいて結果を表示している。そのため、結果上位に利用実績の高い汎用的な部品を表示でき、部品の再利用に関する情報を取得しやすくなる。今後の課題としては、部品グループ化手法の効率化、検索システムの評価および性能向上などが挙げられる。

謝辞 本研究は、科学技術振興事業団計算科学技術活用型特定研究開発推進事業 (ACT-JST) の支援を受けている。

文献

- [1] "SourceForge", <http://sourceforge.net/>
- [2] K. Inoue, R. Yokomori, H. Fujiwara, T. Yamamoto, M. Matsusita and S. Kusumoto: "Component Rank: Relative Significance Rank for Software Component Search", to be appeared in Proceedings of 25th International Conference on Software Engineering (ICSE 2003), Portland, Oregon, 2003.