

G-27

ファイル名を利用したソースコードの機能別グループ化  
Clustering of the source code using the naming convention

○垣谷広輝<sup>1</sup>, 佐藤健広<sup>1</sup>, 田邊純輝<sup>1</sup>, 平山雅之<sup>2</sup>

\*Kouki Kakiya<sup>1</sup>, Katsuhiko Sato<sup>1</sup>, Junki Tanabe<sup>1</sup>, Masayuki Hirayama<sup>2</sup>

Abstract: In order to develop high quality software, improving for software review is highly expected. Especially, in terms of source code review, it is difficult to confirm an association between source code and functional specification in large size of software. In this research, we focus on the file name of source code and develop a support tool which automatically categorizes source files according to their file name. This paper shows a basic idea of our support tool and its experimental evaluation results.

1. まえがき

近年のソフトウェア開発において、レビュー方法の改善需要が高いという問題がある。その中でも特にソースコードと仕様書間の対応がわかりにくいこと、コードの肥大化に伴いレビューに多くの工数がかかること、などの問題がある。この問題解決のため我々はファイル名情報を利用し、詳細仕様書-ソースコード間の対応を自動的にとるシステムを開発中である。今回この方法を簡易的に評価する実験を行い、その有効性を確認したので、報告する。

2. システムの目的

規模の大きなソースコードに関して、ソースコードと仕様書の対応付けを自動的に行うことにより、肥大化するソースコードレビュー工数を削減する方式を検討している。

現時点ではソースコードファイルのファイル名を利用してソースコードファイルを機能単位でグループ化することで、グループと仕様書に記載された機能を一対一で対応させることを目標としている。ファイル名を利用する理由としてはソースコードファイルのファイル名を命名する際にはそのファイルが属する仕様書上に記載された機能に関するキーワードをつけることが多いため、各ソースコードファイル名に含まれるキーワードをインデックスとして、ソースコードファイルをグループ化することで、ソースコードを機能別にグループ化することが可能になるのではないかと考えたためである。

3. 機能

3. 1. 前提とするファイル名の構造

実際の開発でのファイル名は1つの単語をそのまま

用いるのではなく、単語を省略したり、英大文字やハイフン、アンダーバーなどで区切り、複数の単語を組み合わせて使用していることが多い。この単語同士の境界の文字を区切り文字と定義し、これ元にグループ化を行う (Figure1)。

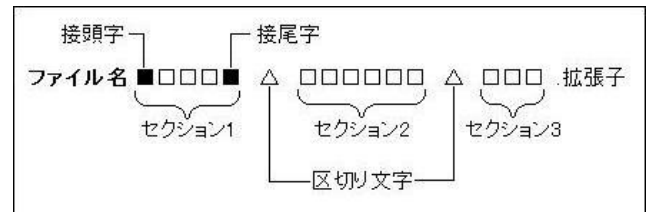


Figure 1. Division of filename

3. 2. グループ化

グループ化は大きく分けて2段階で行う。

(1) 1段階目: セクション文字数によるグループ化

1段階目は拡張子前までの文字数に着目する。区切り文字を境にファイル名を区分しセクションと定義し (Figure1), この各セクションの文字数から分類を行う。Figure2 ではセクションの文字数情報を取り出している。

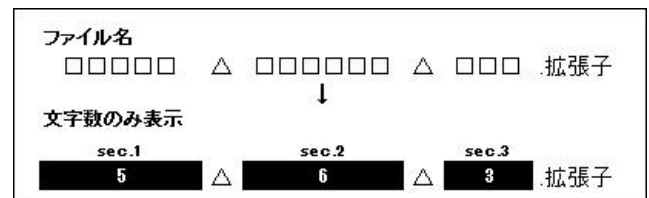


Figure 2. Count of character

(2) 2段階目: 接頭接尾文字を利用したグループ化

2段階目は接頭・接尾など個々の文字に着目する。文字数による分類の場合、異なった単語であっても文字数、あるいは略語の長さが等しければ同じ分類になる。そこで各セクション内の特定箇所の文字を抽出・比較

1: 日大理工・学部・子情 2: 日大理工・教員・子情

することにより、より精度高いグループ化を行う。ここでは接頭・接尾文字を例に取る。Figure3.(a)はセクション1の文字数が5文字のグループである。セクション1は異なる単語であっても文字数が等しいので同じグループとして扱われている。Figure3.(b)のGroupAおよびBでは接頭字と接尾字を比較し、文字が等しいファイル同士をグループとしている。



Figure 3. Clustering

4. 適用評価

4. 1. 評価実験の狙いと方法

提案法を用いることで実際にソースコードと仕様書の対応付けが効率的に行えるかを確認することを目的として簡易実験を行った。実験ではオープンソースソフトウェアである OpenOffice や Teraterm をファイル群分析対象として利用し、このファイル群に含まれるソースコードファイルのファイル名内のキーワードを用いて、機能単位にグループ化できるかどうかを調べた。実験では提案法をベースにした支援ツールを用いてグループ化する場合と支援ツールを用いず人手でグループ化する場合の比較を行った。

4. 2. 実験結果および考察

実験結果 (Figure4) の A,B,C は人手で分類した場合,Program はツールを利用した場合である。Code1~4のいずれかのファイル群においても、提案法を実装したツールを用いて、ソースコードファイルを機能単位の複数のグループに分類することができ、仕様書との対応付けが可能であることを確認できた。また、ツール

を利用した場合と人手でグループ化した場合とでは、分類されたグループ数に有意な差は見られず、各グループに含まれるファイルはほぼ同じだった。グループ化に要した時間については、人手でグループ化する場合よりもツールを利用した場合のほうが早いことが確認できた。一方で、提案法ではファイル名に含まれたキーワードだけでは機能との対応付けが難しいソースコードグループなども検出することが可能であり、これを利用して仕様とソースコード間のトレーサビリティを確認できることが分かった。仕様からコード実装、テストに至るソフトウェア開発プロセスの中で個々の仕様がどのように実装されテストされるかを追いかけていくトレーサビリティチェックは極めて重要視されており、そのための支援ツールとして提案法は有効であると考えられる。

| Code1:ファイル数:41  |     |     |     |         |
|-----------------|-----|-----|-----|---------|
|                 | A   | B   | C   | Program |
| 経過時間(分)         | 18  | 15  | 26  | 0.2以下   |
| グループ数(個)        | 24  | 25  | 26  | 27      |
| Code2:ファイル数:58  |     |     |     |         |
|                 | A   | B   | C   | Program |
| 経過時間(分)         | 16  | 15  | 20  | 0.2以下   |
| グループ数(個)        | 28  | 31  | 26  | 31      |
| Code3:ファイル数:364 |     |     |     |         |
|                 | A   | B   | C   | Program |
| 経過時間(分)         | 63  | 52  | 110 | 0.5以下   |
| グループ数(個)        | 131 | 144 | 102 | 127     |
| Code4:ファイル数:34  |     |     |     |         |
|                 | A   | B   | C   | Program |
| 経過時間(分)         | 20  | 22  | 20  | 0.2以下   |
| グループ数(個)        | 15  | 14  | 26  | 13      |

Figure 4. Experimental result

5. むすび

ソースコードレビューの効率化を目的として、しようとソースコードの対応付けを行うためのソースコードグループ化の方法とその方式を実装したツールの評価を行った。提案法は、ファイル名のみを用いて、簡易な方式で拘束にグループ化を可能としており、機能仕様のトレーサビリティチェックなど、ソフトウェア保守性を評価するための有効な方法となることが確認できた。

6. 参考文献

[1] 三好健吾, 中川雅通:「ソフトウェア構造分析装置」  
 [2] 松村 和哉, 渡部 治朗, 寺内 俊, HE Aiguo:「ソフトウェア可視化手法を用いた初心者向け C 言語教育ツール」