

Title	ソフトウェアライセンスがコピーアンドペーストによる再利用に与える影響の調査
Author(s)	鹿島, 悠; 早瀬, 康裕; 吉田, 則裕 他
Citation	電子情報通信学会技術研究報告. SS, ソフトウェアサイエンス. 2010, 110(227), p. 25-30
Version Type	VoR
URL	https://hdl.handle.net/11094/26668
rights	Copyright © 2010 IEICE
Note	

Osaka University Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

Osaka University

ソフトウェアライセンスがコピーアンドペーストによる再利用に与える 影響の調査

鹿島 悠[†] 早瀬 康裕^{††} 吉田 則裕^{†††} 真鍋 雄貴[†] 井上 克郎[†]

[†] 大阪大学 大学院情報科学研究科

〒 565-0871 大阪府吹田市山田丘 1 番 5 号

^{††} 東洋大学 総合情報学部

〒 350-8585 埼玉県川越市鯨井 2100

^{†††} 奈良先端科学技術大学院大学 情報科学研究科

〒 630-0192 奈良県生駒市高山町 8916-5

E-mail: †{y-kasima,y-manabe,inoue}@ist.osaka-u.ac.jp, ††hayase@toyo.jp, †††yoshida@is.naist.jp

あらまし オープンソースソフトウェア (OSS) のソースコードは、ライセンスに示された条件を満たすことで再利用することが出来る。ここで、OSS のライセンスには多くの種類が存在するため、再利用の際に満たさなければならない条件も様々なものが存在する。そこで我々は、この条件によって、ソースコードが再利用される頻度や、再利用先のソフトウェアの種類などに違いが発生しているのではないかと考えた。ライセンスが再利用に与える影響を調査するため、ソースコードの最も基本的な再利用の方法であるコピーアンドペーストを、コード片が含まれるファイルのライセンスで分類する。本論文では、予備的な調査として小規模なソースコード集合に対して、手動による予備的な調査を行なった。その結果、ライセンスによって、コピーアンドペーストの頻度が異なることや、コピーされたコード片を共有するソフトウェアの組には、ライセンスの点で特徴があることが分かった。

キーワード ソフトウェアライセンス, オープンソースソフトウェア, 再利用, コピーアンドペースト

An Investigation of Copy-and-paste Reuse between Source Files Distributed under Different Licenses

Yu KASHIMA[†], Yasuhiro HAYASE^{††}, Norihiro YOSHIDA^{†††}, Yuki MANABE[†], and Katsuro
INOUE[†]

[†] Graduate School of Information Science and Technology, Osaka University

1-5, Yamadaoka, Suita, Osaka, 565-0871, Japan

^{††} Faculty of Information Sciences and Arts, Toyo University

2100 Kujirai, Kawagoe, Saitama, 350-8585, Japan

^{†††} Graduate School of Information Science, Nara Institute of Science and Technology

8916-5 Takayama-cho, Ikoma, Nara, 630-0192 Japan

E-mail: †{y-kasima,y-manabe,inoue}@ist.osaka-u.ac.jp, ††hayase@toyo.jp, †††yoshida@is.naist.jp

Abstract Source code of open-source software (OSS) can be reused if a developer observe requirements of its license. There are many various licenses, therefore there are many various requirements to reuse source code. Then, we presume that this condition influences frequency of reusing source code and variety of software. To analyze an influence of license to reuse, we make a classification of copy-and-paste —this is the most basic method of reusing source code— in the basis of a license of a file which contains code snippet. In this paper, we performed preliminary experimental study that analyzes a small-scale set of source code by hand. The result shows that there is a difference of frequency of copy-and-paste between each license, and a couple of software which shares code snippet copy-and-pasted is characterized by its license.

Key words Software License, Open Source Software, Reuse, Copy and Paste

1. はじめに

オープンソースソフトウェア (OSS) は、ソースコードを公開し、自由な再頒布や改変が行えるようにしたソフトウェアである。現在、多くの OSS が公開されており、利用可能なソースコードが増えている [1]。

ソフトウェアを利用するには、そのソフトウェアに付随するライセンスに従わなければならない。OSS において、ソフトウェアやそのソースコードの利用方法についての開発者の意図は様々である。具体的な開発者の意図としては、ソフトウェアの利用で生じた不具合について責任を負わないという意図、ソフトウェアの二次利用を行う際は二次著作物も OSS にしたいという意図が挙げられる。そのため、OSS では開発者の意図に合わせた多くのライセンスが存在している。

ソフトウェア開発を効率的に行う方法として、既存のソフトウェアを利用して新しいソフトウェアの開発を行うソフトウェアの再利用がある。その中でもソースコードの再利用を行う際に用いられている方法の一つが、ソースコードのコピーアンドペースト (C&P) である。ソフトウェア再利用を行うことにより、ソフトウェアの信頼性や生産性が向上すると言われている。OSS でも再利用は行われており、他の OSS の全部または一部を再利用していることがある。

ソフトウェアの再利用は、ソフトウェアの利用の一形態であり、再利用を行う際には、再利用されるソフトウェアのライセンスに従わなければならない。さらに、再利用したソフトウェアを組み込む先のライセンスにも抵触しないようにしなければならない。

OSS で利用されるライセンスは、C&P による再利用を行う際、元のライセンスとは別のライセンスで頒布する際の条件が、頒布が可能かどうかも含めて大きく異なっている [2]。例えば、例えば GNU General Public License (GPL) [3] では、GPL のソースコードを C&P して再利用した場合、GPL のソースコードを再利用して作られたソフトウェアのライセンスも GPL にしなければならない。別の例として、BSD ライセンス [4] では、著作権、ライセンス条文、免責条項の明記さえ行えば自由に C&P を行ってよい。

しかし、再利用するソフトウェアのライセンスの条件を満たすために、作成するソフトウェアのライセンスを変えるのは困難な場合がある。そのため、再利用するソフトウェアのライセンスを制限するライセンスとなっているソフトウェアが再利用される頻度は少ないと考えられる。しかし、現状我々が確認した限りでは、C&P による再利用を、ライセンスの観点で分析した研究は行われていない。そのため、ライセンスが再利用に実際に与えている影響は未だ不明である。

そこで本研究では、実際のオープンソースソフトウェアに対し、C&P による再利用をライセンスの観点で調査するための予備的実験を行い、調査方法が妥当かどうかを検証した。調査対象としたソースコードは、Debian GNU/Linux [5] のパッケージ管理システムで管理されている main パッケージに含まれる Java で記述されたソースファイル集合である。また、調査対象とし

たライセンスは、BSD 3 項ライセンス (BSD3) [4]、GPL Version 2 or later (GPLv2+)、Apache License 2.0 (Apachev2) [6] である。

予備的実験の結果、BSD3 や Apachev2 は、GPLv2+ に比して C&P による再利用が積極的に行われていたことを確認した。また、Apachev2 や GPLv2+ では、特定のライセンスのソースコードとの C&P による再利用が頻繁に行われていたことを確認した。

以上の結果を得て我々は、今回の予備的実験の手法を用いることで C&P による再利用をライセンスの観点で分析することに成功したと考え、本手法の妥当性を確認した。

以降、2. 節では本研究の背景を、3. 節ではライセンスが再利用に対して与える影響を、4. 節では実験の方法とその結果を、5. 節では実験の考察を、6. 節では実験と考察の妥当性を、7. 節ではまとめと今後の課題を述べる。

2. 背景

2.1 ライセンス

OSS では様々なライセンスが利用されている。また、そのライセンスソフトウェアを利用する際の条件も様々である。

OSS で利用されているライセンスのうち、代表的なものとして Apachev2、GPL、BSD3 が挙げられる。これらのライセンスはそれぞれソフトウェアを利用する際の条件が異なっている。

Apachev2 のソフトウェアを利用し、二次著作物を頒布する際には、著作権、免責条項、特許権、商標権に関する記述はそのまま保持されなければならない。改変が行われた場合は、元のソフトウェアとの変更点を示さなければならない。

GPL のソフトウェアを利用し、二次著作物を頒布する際には、元のソフトウェアとの変更点を示し、かつ、二次著作物のライセンスも GPL にしなければならない。

BSD3 のソフトウェアを利用し、二次著作物を頒布する際には、著作権、免責条項、ライセンス条文の明記が必要である。また、許可無く元のソフトウェアの著作者名や組織名を、ソフトウェアの宣伝に利用することを禁じている。

そして、現在ソースコードのライセンスを示す方法として用いられているのは主に次に挙げる 2 つの方法である。

一つは、ソースコードの先頭にそのソースコードのライセンスを示す文章をコメントとして記述する方法である。

もう一つは、ライセンスを示すテキストファイルをソースコードと共に配布して、ソフトウェアの全体または一部のライセンスをまとめて示す方法である。

2.2 再利用とライセンス

信頼性の高いソフトウェアを効率良く生産する方法として、ソフトウェアの再利用が行われている。ソフトウェアの再利用とは、既存のソースコードやドキュメントなどソフトウェアの開発過程で生産される成果物を同一システム内や他のシステムで利用することである。

OSS ではバイナリだけでなく、ソースコードも配布されている。そのため、ソースコードを再利用することが可能である。

ソースコードを再利用する際に用いられる方法の一つが C&P

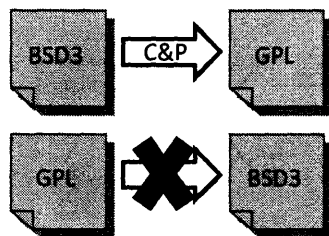


図1 ライセンスによる C&P の可不可

である。C&P では、再利用するソースコードの全部または一部を複製し、必要があればそのソースコードを改変し、開発するソフトウェアのソースコードとして利用する [2], [7]。

C&P によるソースコードの再利用を行う際には、まず、再利用されるソフトウェアのライセンスに従わなければならない。同時に、再利用を行って開発するソフトウェアのライセンスにも従わなくてはならない。

そのため、再利用される側のソフトウェアと、再利用を行い開発しようとしているソフトウェアのライセンスが異なっている場合、どちらかあるいは両方のライセンスの条件を満たせない場合がある。例えば GPL のソフトウェアのソースコードを C&P して再利用する場合、再利用を行って開発しようとしているソフトウェアのライセンスが、GPL 以外のライセンスであった場合、開発しようとしているソフトウェアのライセンスを GPL にしない限りは GPL のライセンスに違反していることになる。

ライセンスしだいで、C&P ができる場合とできない場合があることを示した図を図 1 に示す。BSD のソースコードから GPL のソースコードへの C&P は問題ないが、GPL のソースコードから BSD のソースコードへと C&P した場合、そのままでは、GPL のライセンスに違反している。

3. 再利用におけるソフトウェアライセンスの影響

開発者が再利用を行いたい場合でも、実際に再利用が行えるかどうかは、再利用したいソースコードのライセンスに大きく影響される。2.2 節で述べた通り、ソースコードの再利用を行う際には再利用されるソースコードのライセンスと再利用を行って開発するソフトウェアのライセンスに従わなくてはならない。そのため、再利用されるソフトウェアのライセンスが、GPL のように開発するソースコードのライセンスを制限するライセンスである場合は、再利用を行いたい場合でも、ライセンスの条件を満たすのが難しい場合がある。逆に、ライセンスが BSD であるソフトウェアは、ライセンスの条件を満たすのが容易であるため、再利用を行う際の障害は少ない。つまり、再利用する際の条件を満たすのが容易なライセンスとそうではないライセンスがある。

よって、再利用ための条件を満たすのが容易なライセンスが使用されているソースコードは、様々なライセンスのソースコードで再利用されていると予想される。また、再利用される回数も、再利用のための条件を満たすのが難しいライセンスを

使用しているソースコードに比べて多いと予想される。

しかし、ソースコードの再利用を、ライセンスの観点で調査した研究は、我々が調べた限りでは確認されておらず、ライセンスが再利用に与えている影響は未だ分かっていない。よって、本研究では C&P によるソースコードの再利用をライセンスの観点で調査する。

4. 実験

本実験の目的は、OSS において、ライセンスが C&P による再利用に実際にどのような影響を与えているかの調査である。具体的には、OSS で再利用が行われる際、どのようなライセンスどうしで C&P が行われやすいか、またそれは、ライセンスごとに異なっているのかどうかを調査する。

以上を目的とした調査を実際の OSS を対象に行い、実際の OSS で C&P がどのようにライセンスと関連しているのかを調査する。ただし、今回の実験では、まず実験の手法が正しいかどうかを調べるために、小規模な OSS のソースファイル集合に対して実験を行った。

また、同じアプリケーション内で行われている C&P は、たとえライセンスが違うファイルどうしでの C&P であったとしても、ライセンスの問題は必ず解決されているはずであり、ライセンスが与えている影響は少ないと考えられる。よって、今回の実験では、互いに異なるソフトウェアに属するソースコードの C&P のみに注目する。

4.1 実験対象

本実験では、調査の対象として Debian GNU/Linux のパッケージ管理システムにて管理されているパッケージのソースファイル集合を対象とした。その理由は以下の通りである。

- Debian GNU/Linux のパッケージ管理システムで管理されているパッケージのソースファイルのライセンスには、様々なライセンスが使用されている点
- 含まれているパッケージ数、ソースファイルの数が大きく、規模の大きいソースファイル集合である点
- 保守・管理が続いており、多数のユーザーもいるパッケージの集合である点

そして、パッケージ管理システムが管理しているパッケージの中でも、Debian Project が直接管理しているパッケージの集合を対象にした。さらにその中でも、Java で記述されたソースコードが含まれているパッケージを対象とした。

ただし、Java で記述されたソースコードを含むパッケージの中には、同じパッケージでバージョンが異なっているだけというパッケージ^(注1)がいくつか含まれており、それらのパッケージに含まれるソースファイルは調査対象から除外した。

今回実験対象としたソースファイル集合のパッケージ数は 452、ファイル数は 77452、行数は 8530896 である。

4.2 実験方法

本実験の手順の概要を図 2 に示す。

本実験では、まず各ソースファイル中のライセンスを示す記

(注1) : antlr, asm, asm2, db4.2, db4.5, junit, tomcat5.5

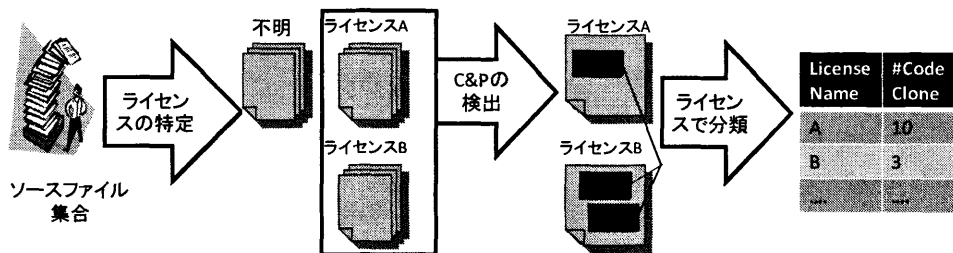


図 2 実験手順の概要

述が何のライセンスを示しているのかを読み取ることで、各ソースファイルのライセンスの特定を行う。ライセンスの特定には Ninka [8] を利用した。Ninka はソースファイルのコメント中のライセンスを示す記述を解析し、その記述がどのライセンスを指しているかを特定するツールである。

次に C&P により生成されたクローンセットの集合を抽出する。クローンセットとは、クローン関係の同値類である。コード片 α とコード片 β が類似したコード片であるとき、 α と β はクローン関係にあるという。クローン関係は同値関係であるため、同値類を作成することができる。また、類似したコード片のことをコードクローンと呼ぶ。クローンセットの抽出を行う際には、異なるパッケージに所属するファイルどうしで C&P されているコード片の集合を抽出する。クローンセットの集合の抽出の具体的な方法は 4.3 節にて述べる。

そして、クローンセット中にあるライセンス A のコード片が含まれているクローンセットを抽出し、クローンセット中に含まれているコード片を、各コード片のライセンスごとに分類して数え上げることで、A のコードがどのライセンスのコードとクローン関係にあるのかを調べる。これにより、A のコードがどのライセンスにコードに C&P された、あるいはどのライセンスのコードを C&P したのかを調べる。ただし、今回の実験では C&P のコピー元とコピー先を判別しないため、C&P されたのかしたのかは、判別できない。

全てのライセンスに対して上記の操作を行うのは手間がかかりすぎるため、本実験では、実験対象のソースファイル集合において、多くのファイルで使われており、かつ、再利用の際の条件が互いに異なっている複数のライセンスを調査対象とする。表 1 に、実験対象のソースファイル集合における各ソースファイルをライセンスで分類した結果、ファイル数が多い順に上位 10 件を示す。

なお、表中のライセンス名は省略名称を使っており、主な省略名称は表 2 にて示している。バージョン番号を示す場合は省略名称の後ろ「v」を付け、その後ろにバージョン番号を付ける。また、特定のバージョン or Later と記してあるライセンスの場合は、バージョン番号の後ろに「+」を付けて表わす。また、複数のライセンスが選択可能になっている場合は、ライセンス名を「,」で続けて表記する。

最終的に我々は、Apachev2, GPLv2+, BSD3 を調査対象とした。これらのライセンスは実験対象にて多く使われており、かつ再利用の際の条件はそれぞれ異なっていると我々は判

表 1 ソースファイル全体のライセンスの分布

License Name	#File
Apachev2	16350
GPLv2+	8160
LesserGPLv2.1+	6534
GPLnoVersion,GPLv2+,LinkException	5887
GPLv2	3222
SeeFile	2786
BSD3	2181
GPLv2,ClassPathException	1498
ライセンスを示す記述が無いファイル	15813
解析に失敗したファイル	6862

表 2 ライセンスの省略記法

Abbreviation	Name
Apache	Apache Public License
BSD3	Original BSD minus advertisement clause
CPL	Common Public License
ClassPathException	GNU Classpath License
GPL	General Public License
LesserGPL	Lesser General Public License
LibraryGPL	Library General Public License
LinkException	GPL linking exception
MITX11noNotice	MIT License/X11License
MPL	Mozilla Public License
MX4JLicense	MS4J License
SeeFile	File Points to another where the its license is
publicDomain	Public Domain
subversion	Subversion License

断した。

4.3 C&P の検出方法

C&P の検出には、CCFinder [9], [10] を利用する。CCFinder は、コードクローン検出ツールであり、コードクローンの検出を行うことで C&P にされたコード片の検出を行う。

CCFinder で C&P の検出を行う際には、実験対象のソースファイル集合から、識別子名を無視して、クローンセットの検出を行う。識別子名を無視するのは、C&P を行った後識別子名だけを変えたコード片をクローンセットとして検出するためである。

さらに、CCFinder が出力する LNR という値を用いて、検出されたクローンセットのうち、C&P により生成されたコード片を含むクローンセットを抽出する。LNR はコード片に含まれる非繰り返し要素のトークン数である。LNR が小さいコード片は、変数宣言、代入文、アクセサ宣言、メソッド呼出などが連続しているだけのコード片である可能性がある。逆に LNR が大きいコード片は C&P により生成された可能性が、LNR が低いものと比べて高い。そこで、本実験では、クローンセット

表 3 BSD3 のファイルとクローン関係にあるコード片の分布

License Name	#Code
BSD3	613
GPLv2+	20
Apachev2	16
LesserGPLv2+	14
GPLv2,ClassPathException	1
LesserGPLv2.1+	1

表 4 Apachev2 のファイルとクローン関係にあるコード片の分布

License Name	#Code
Apachev2	1533
Apachev1.1	316
LesserGPLv2.1+	42
MPLv1.1	33
BSD3	29
MX4JLicensev1	16
GPLv2+	4
LibraryGPLv2+	3
MPLv1.0	2
MITX11noNotice	2
publicDomain	1
subversion+	1
EPLv1	1

表 5 GPLv2+ のファイルとクローン関係にあるコード片の分布

License Name	#Code
GPLv2+	268
GPLnoVersion,GPLv2+,LinkException	225
BSD3	28
LibraryGPLv2+	20
Apachev2	4
LesserGPLv2.1+	4

表 6 コード片の総数とファイル数

	#Code	#File	#Code/#File
BSD3	665	2181	0.304906
Apachev2	1983	16350	0.121284
GPLv2+	549	8160	0.067279

中のコード片の LNR の平均値が 50 以上のクローンセットを C&P により生成されたクローンセットと判断した。

4.4 実験結果

BSD3 のコードが含まれるクローンセットに含まれるコード片をライセンスごとに分類し、数え上げた表を表 3 に示す。Apachev2 の結果は表 4 に、GPLv2+ の結果は表 5 に示す。

また、各ライセンスの分析結果から、コード片の総数を算出し、実験対象中のライセンスのファイル数との比較した表を表 6 に示す。表より、ファイル数に比して C&P の関係にあるコード片の総数が多い順に、BSD3, Apachev2, GPLv2+ となることが分かる。

全てのライセンスの結果に共通する特徴として、各ライセンス自身のコード片が多数含まれていたことが挙げられる。BSD3

の結果では、BSD3 のコード片が、コード片全体の 92% を占めていた。同様に Apachev2 では、Apachev2 のコード片が全体の 77% を占め、GPLv2+ では GPLv2+ のコード片が全体の 48% を占めていた。

また、Apachev2 の結果には、Apachev1.1 がコード片全体の 16% を占めており、Apachev2 に続いて多く、Apachev2 以外のライセンスと比べて抜きん出て多い。

同様に GPLv2+ の結果では、「GPLnoVersion, GPLv2+, LinkException」がコード片全体の 41% を占めており、これも GPLv2+ 以外と比べて抜きん出ている。

クローン関係にあるライセンスの種類数という観点で評価すると、種類数が一番多いのは Apachev2 であり、Apachev2 自身を含めて 13 種類のライセンスと C&P の関係にあった。BSD3 と GPLv2+ は、共に 6 種類のライセンスとの C&P の関係が認められた。

5. 考 察

Apachev2 の分析結果で、Apachev1.1 とクローン関係にあるコード片が多く見られた。その理由として、Apachev1.1 は Apachev2 の旧版であり、Apachev2 へライセンスを変更する作業が進められている、ということが考えられる。

また、GPLv2+ は表 6 の #Code/#File が最も少なく、BSD3 や Apachev2 と比して C&P による再利用が積極的に行われていないと考えられる。これは、GPLv2+ は、Apachev2 や BSD3 や GPL 系列のソースコードを C&P してはいても、GPLv2+ のソースコードが C&P されることはあまり無かったためだと考えられる。この理由を以下に示す。

- GPLv2 または GPLv3 以外のライセンスのソースコードに、GPLv2+ のソースコードの C&P するのは GPLv2+ のライセンスに違反する
- Apachev2 や BSD3 のソースコードは GPLv2+ のソースコードに C&P してもよい
- LesserGPL や「GPLnoVersion, GPLv2+, LinkException」等の GPL 系列のライセンスは、GPL にライセンスを変更して C&P してもよい

逆に BSD3 や Apachev2 では、GPLv2+ と比べて C&P による再利用が積極的に行われていた。この理由は、BSD3 や Apachev2 のソースコードを C&P して再利用する際の条件を満たすのは容易だからだと推測される。実際、Apachev2 では GPLv2+ と比して多くの種類のライセンスと C&P の関係にあり、他のライセンスのソースコードに積極的に C&P されたことが予想される。

全ライセンスに共通する特徴としては、4.4 節で、各ライセンスの分析結果の中で最も多かったのは、分析対象にしたライセンス自身のコード片であったということを挙げた。これは、開発団体内で利用するライセンスを統一している場合に、同じ開発団体が作成したソースコードを C&P することが多かったからではないかと推測される。

なお、本実験は予備の実験であり、実験対象を大規模にして引き続き実験を行う予定である。その際、C&P の検出につい

ては、CCFinder は入力ソースファイル集合を分割して与え、その後出力を統合することが可能であるため、大規模でも適用可能である。また、ライセンスの特定については、Ninka は一つのファイルを入力としたツールであり、実験対象が大規模であっても一ファイルずつ処理すればよいので大規模でも適用可能である。よって、本手法はより大規模な実験対象でも適用可能であると考えている。

6. 妥当性

本実験では、C&P の検出に CCFinder を用いて、非繰り返し要素が平均 50 トークンより多く出現しているコード片の集合を C&P で生成されたコード片の集合とみなしている。しかし、C&P で無くても、50 トークンより多い非繰り返し要素が別々のソースファイルに出現する場合も無いとは言えず、CCFinder により検出された C&P の全てが、本当に C&P により生成されたとは言いきれない。とはいえ、C&P 以外の理由で生成されたコードクローンが多くを占めているとも考えづらく、本実験で CCFinder を用いたのは妥当であると考えている。

本実験では、各ソースファイルのライセンスの特定に Ninka を用いた。ライセンスを示す記述がソースファイルにある場合に Ninka が正しくライセンスを特定する、再現率が 82%、適合率が 96% であり [8]、かつ、実際に多くのファイルのライセンスを特定できたため、Ninka を用いたのは妥当であると考えている。しかし、ソースファイル中にライセンスを示す記述が無い場合や、使用されているライセンスが Ninka に登録されていない場合には、Ninka ではライセンスの特定を行うことはできない。今回の実験では、Ninka がライセンスを特定できなかった場合は C&P の検出の対象から外しており、ライセンスが特定できなかったソースファイルが実験結果に影響を与えないようにしている。

今回実験対象としたのは、Debian GNU/Linux のパッケージ管理システムで管理されているソースファイル集合の中の極一部のソースファイルでしかなく、標本誤差を多分に含んでいる可能性がある。そのため、本実験の結果を OSS 一般に適用するのは難しい。しかし、今回の実験対象とソースファイルのライセンスの分布が同様である対象には、本実験の結果を適用できると考えている。

7. まとめと今後の課題

本研究では、ライセンスが C&P による再利用にどのような影響を与えているか、その調査方法が妥当かどうかを確かめる予備的実験を行った。実験では、Debian GNU/Linux のパッケージ管理システムで管理されているソフトウェアの一部を対象に、C&P をライセンスの観点で分析した。その際、BSD3 との C&P、Apache2 との C&P、GPLv2+ との C&P を分析した。

その結果、C&P の多くは同じライセンス、あるいは同じ団体が規定したライセンスのファイルどうしで行われていることを確認した。また、Apache2 は他の 2 つのライセンスと比べ、多くのライセンスのソースコードと C&P の関係にあることが

分かった。逆に、GPLv2+ は同じ団体が規定したライセンスを除けば、少数のライセンスのソースコードとしか C&P の関係に無かった。

本実験は、あくまで予備的実験であり、本研究の手法を Debian パッケージのソースファイル集合全体に適用し、OSS 全体でライセンスが C&P による再利用にどのような調査を与えているかを調査する予定である。

また、今回の実験では、C&P により生成されたクローンセットに焦点を当てただけで、どのコード片が、C&P される元のコード片となっていたかは解析していない。そのため、どのライセンスから、どのライセンスへと C&P されたのかは不明であり、これを分析するのを今後の課題としたい。

そして、考察で述べた、同じ開発団体内で C&P が行われたのでは無いか、という予想の検証も、今後の課題としたい。

謝 辞

本研究は、日本学術振興会科学研究費補助金基盤研究 (A) (課題番号:21240002) および文部科学省科学研究費補助金若手研究 (B) (課題番号:21700031) および日本学術振興会科学研究費補助金研究活動スタート支援 (課題番号:22800040) の助成を得た。

文 献

- [1] W. Scacchi: "Free/open source software development: recent research results and emerging opportunities", ESEC/FSE 2007, pp. 459–468 (2007).
- [2] M. Ruffin and C. Ebert: "Using open source software in product development: A primer", IEEE Software, **21**, 1, pp. 82–86 (2004).
- [3] Free Software Foundation: "GNU general public license". <http://www.gnu.org/licenses/gpl.html>.
- [4] Open Source Initiative: "The BSD license". <http://www.opensource.org/licenses/bsd-license.php>.
- [5] Debian Project: "Debian gnu/linux". <http://www.debian.org/>.
- [6] Apache Software Foundation: "Apache license, version 2.0". <http://www.apache.org/licenses/LICENSE-2.0>.
- [7] J. Li, R. Conradi, C. Bunse, M. Torchiano, O. P. N. Slyngstad and M. Morisio: "Development with off-the-shelf components: 10 facts", IEEE Software, **26**, pp. 80–87 (2009).
- [8] D. M. German, Y. Manabe and K. Inoue: "A sentence-matching method for automatic license identification of source code files", ASE 2010, pp. 437–446 (2010).
- [9] T. Kamiya, S. Kusumoto and K. Inoue: "CCFinder: A multilingual token-based code clone detection system for large scale source code", IEEE Transactions on Software Engineering, **28**, pp. 654–670 (2002).
- [10] 肥後 芳樹, 吉田 則裕, 楠本 真二, 井上克郎: "産学連携に基づいたコードクローン可視化手法の改良と実装 (情報システム開発 < 特集 > 産学連携論文)", 情報処理学会論文誌, **48**, 2, pp. 811–822 (2007).