

いくつかのOSSに対する細粒度分析を用いた著作権表示の調査

田邊 傑士[†] 井上 克郎[†]

Investigation of OSS Copyright Notices Using Fine-Grained Analysis

Taketo TANABE[†] and Katsuro INOUE[†]

あらまし 著作権表示とはソースコードの著作権者を示したファイル先頭部分のコメントである。OSSの場合、多くの人ソースコードを作成・編集することが可能であり、全ての人に著作権者を示すモチベーションがあるとは限らない。そこで本研究ではどの程度、著作権表示が記述されているのかを調べるため、ソースコードを行単位で著作権表示の有無を定義した上で、その割合を調査した。結果としてどのOSSにも著作権表示がない行が存在していること、著作者が作成した行数の割合と著作権表示を記述する割合に関係があることがわかった。

キーワード 著作権表示, 著作権者, OSS, git blame

1. ま え が き

OSS(Open Source Software)は作成者がソースコードを無償で公開しているソフトウェアである[1]。OSSでは誰でも公開されているソースコードの開発に加わることができる。例えば、有名なOSSの1つであるLinuxカーネルにおいては2019年末までに21,074の異なる人物が開発に携わっている[2]。また、著作権表示とはソースコードの著作権を持つ人や組織(著作権者)を示したファイル先頭部分のコメントである。通常、著作権とは著作物を作成した時点で発生するものであるため、著作権の取得のために著作権表示を記述する必要はない[3]。よって様々な人や組織が開発に携わることがあるOSSにおいて、全ての開発者が著作権表示を記しているとは考えづらい。そこで、OSSにおける著作権表示の実態を調べるために本研究を行った。

本研究ではOSSにおける著作権表示の実態を調べるために独自の基準を導入した。まず、調査対象のソースコードを行単位の粒度で分割し、各行に対して著作権表示の有無を独自に定義した。そして、著作権表示があると判断された行数がどの程度あったかを調べ、これを著作権表示がどの程度記述されているのかの基準とした。

この論文は次のように構成している。はじめに2.節

でOSSと著作権に関する簡単な背景説明を行った後、3.節で調査における定義と調査方法について説明する。4.節では、実施した調査の結果を示す。5.節では調査結果に対する考察について述べる。そして6.節では調査結果に対する妥当性の脅威とその影響について述べる。最後に7.節で論文を締めくくり、今後の研究の方向性について概説する。

2. 背 景

この節では調査内容を理解する上で前提となる著作権とOSSについて説明し、OSSの著作権表示とその問題点を概説する。

2.1 著 作 権

著作権と呼ばれる権利は支分権という権利の集合である[4]。支分権には主に著作物を公表する権利、著作物を頒布する権利、著作物を複製する権利、著作物とその名前を改変されない権利などがある。著作権に関する法律の内容は国によってやや異なるものの、その多くは世界知的所有権機関(WIPO)によって設定された共通のガイドラインに基づいており、大きな違いは無い[5]。そして著作権はTRIPS協定などの国際協定によってソースコードに対しても適用する事が定められている[6]。

2.2 職 務 著 作

著作権には職務著作と呼ばれる概念が存在する。これは職務の範囲で作られた作品の著作権は作成者が所属する組織に帰属するというものである。この概念は

[†]大阪大学

```

/*
 * arch/arm/include/asm/glue-proc.h
 *
 * Copyright (C) 1997-1999 Russell King
 * Copyright (C) 2000 Deep Blue Solutions Ltd
 */

```

図1 著作権表示の例

必ずしも全ての国で認められているわけではないが日本をはじめ多くの国で適用されている [7].

2.3 OSS

OSS(Open Source Software) は作成者がソースコードを無償で公開しているソフトウェアである [1]. 一般に作成者は利用者に特定のライセンスの下でソフトウェアの自由な利用を認めている. だれでも自由に編集や再配布が可能という特徴を持つため OSS には通常, 複数人の開発者が存在する.

2.4 OSS の著作権表示

世界の多くの国で, ソースコードは著作権法によって保護されている [6]. それは OSS に対しても同様である. OSS の著作権者は著作権を放棄しているわけではなく, ソフトウェアにライセンスを付与し, 利用者がライセンスの条項を遵守することを条件に本来は著作権者のみがかつ支分権の一部の利用を許可しているに過ぎない. しかし, OSS は誰でも開発に携われるため, 開発者以外のソースコードの閲覧者が著作権者を特定することは容易ではない. そこで著作権者を知る有効な方法の1つが著作権表示を見ることである. 著作権表示とはソースコードの著作権者を示したファイル先頭部分のコメントである. これによって開発者は自らの著作権を閲覧者に主張し, 閲覧者は著作権者を知ることができる. 図1は実際に Linux のファイルの1つに記述されている著作権表示の例である. これは arch/arm/include/asm/glue-proc.h のソースコードの著作権者が Russel King 氏と Deep Blue Solutions 社であることを示している.

一方で, 著作権表示には2つの問題が存在する.

1つはすべての著作権者が著作権表示を記述しているとは限らないという点である.

2つは著作権表示に示されている著作権者がファイルのソースコード全体のうち, どれだけのソースコードの著作権を持っているかがわからないということである. 過去には Linux カーネルの開発者の1人である Patrick McHardy が Linux カーネルにおける自身の著作権を主張して Linux カーネルを利用した企業を訴え

て金銭を要求した事例があり, その際には彼が実際にどれだけのソースコードの著作権を持っているのかが争点の1つとなった [8].

2.5 関連研究

OSS において人とソースコードを結びつけ, 誰がどの程度のソースコードを作成したかを調べる研究が Daniel German らによって行われている [9]. また, OSS における著作権表示の不一致を調べる研究が Qiu らによって行われている [10].

本研究では著作権表示とソースコードを結びつけ, ソースコードに対する著作権表示の有無の程度を調べる.

3. 調査方法

この節ではまず, 調査に使用したツールについて説明する. 次に本研究において独自に定義した著作権がある行についてそのように定義した理由とともに説明する. その上で, 調査対象となった OSS を列挙し, 調査手順を示す.

3.1 調査に使用したツール

3.1.1 git

git とは無料の分散型バージョン管理システムである [11]. バージョン管理システムとはプロジェクトに対する過去の変更や状態を蓄積したものである. この蓄積されたものをリポジトリや git リポジトリと呼び, 変更をコミット (commit) と呼ぶ. また, git には編集履歴を利用した git コマンドと呼ばれる様々なコマンドが用意されている. 本研究では主に git blame コマンドを使用した. これは引数にファイル名を入れることでファイルの各行に対して, その行を作成したコミット, 作成者, 作成者のメールアドレスなどの情報を一覧にして出力するコマンドである.

3.1.2 GitHub

GitHub とは, git リポジトリをインターネット上に公開し他の開発者と git のシステムで共同開発することを可能にしたプラットフォームである [12]. 利用者は GitHub 上の任意の git リポジトリをローカルリポジトリにコピーすることで, ローカルリポジトリにコピーした git リポジトリを git コマンドで操作することが可能となる. 多くの OSS プロジェクトや個人がソースコードやバイナリを GitHub に公開している. 本調査では GitHub から調査対象とした OSS の git リポジトリを取得した.

3.2 著作権表示がある行の定義

(1) ソースコードの各行に対して (2) の場合を著作権表示がある、それ以外の場合を著作権表示がないという。

(2) その行の著作権者がその行を含むファイルの著作権表示に示されている著作権者の1つと一致する。

しかし、実際にはソースコードの著作権者を正確に得ることと、ファイルの著作権表示に示されている著作権者を機械的に得ることは困難である。そこで本研究では以下に述べるような仮定のもと、それらの情報を `git blame` から得られる情報で代用する。

3.2.1 定義上の仮定

`git blame` では各行の作成者のメールアドレスが得られる。本研究ではこのメールアドレスを著作権者の情報として利用する。また、一般にメールには個人的に利用できるフリーメールと組織から支給されるメールがある。そこで本研究では以下の3項目を仮定する。仮定の3項目目は職務著作の考えに基づいている。

- メールアドレスから個人を識別できる
- フリーメールの使用者は個人開発者であり、それ以外の者は職務として開発を行っている
- 著作権表示には著作権表示の作成者が個人開発者の場合は作成者の名前が、それ以外の場合は作成者が所属する組織の名前が著作権者として示されている

フリーメールかどうかはメールアドレスのドメインから判断できる。ドメインがフリーメールのドメインであることは T. Brian Jones によって現時点で 3,782 のフリーメールのドメインがまとめられたリストに含まれていることとした¹。

3.2.2 (2) の条件の代用と実際の使用例

前節までを踏まえ、本研究における (2) の条件は以下の (3), (4) のいずれかを満たすことを言う。

(3) その行の作成者が個人開発者であるとき、その行の作成者のメールアドレスと著作権表示の作成者の1人のメールアドレスが一致する

(4) その行の作成者が職務として開発を行っているとき、その行の作成者のメールアドレスのドメインと著作権表示の作成者の1人のメールアドレスのドメインが一致する

(3) の例と (4) の例を次に示す。ここでは `free-mail.com` はフリーメールのドメイン、`company.com` は組織のドメインとする。例1の場合、15行目の作成

OSS 名	バージョン	対象 ファイル数	対象行数
linux	5.8	42,691	17,911,842
FFmpeg	4.4	3,777	1,466,870
libuv	1.41.0	345	101,423
obs-studio	26.1.2	1,359	414,333
redis	2.04	37	18,895

表1 調査対象の OSS 一覧

者は個人開発者であり、その作成者のメールアドレスが著作権表示の作成者のメールアドレスと一致するため (3) の条件を満たし、15行目は著作権表示があると判断される。例2の場合、15行目の作成者は職務として開発を行っており、その作成者のメールアドレスのドメインが著作権表示の作成者のメールアドレスのドメインと一致するため (4) の条件を満たし、15行目は著作権表示があると判断される。

例1:個人開発者

```
1. [著作権表示:
    作成者 A(tanabe@free-mail.com)]
    ...
15.[ソースコード:
    作成者 A(tanabe@free-mail.com)]
    ...
```

例2:職務として開発を行っている人

```
1. [著作権表示:作成者 B(inoue@company.com)]
    ...
15.[ソースコード:作成者 C(qiu@company.com)]
    ...
```

3.3 調査対象

調査対象とした OSS の一覧を表1に示す。プログラムの都合上、本研究では C または C++ を主に使用している OSS を調査対象とした。また、各 OSS のバージョンは調査開始時点での最新バージョンである。本研究では対象ファイルは各 OSS を構成するプログラムに関するファイル (C または C++) のファイルのみとなっている。

3.4 調査手順

本研究では著作権表示がどの程度記述されているかを調べるにあたってソースコードの各行に対して著作権表示の有無を定義し、その行数の割合を調べた。調査対象の OSS は Github から git リポジトリを取得している。

(注1) : `free_email_provider_domains` <https://gist.github.com/tbrianjones/5992856/>

OSS 名	著作権表示がある 行の割合	著作権表示がない 行の割合
linux	75%	25%
FFmpeg	59%	41%
libuv	48%	52%
obs-studio	63%	37%
redis	82%	18%

表 2 OSS 全体の著作権表示の有無

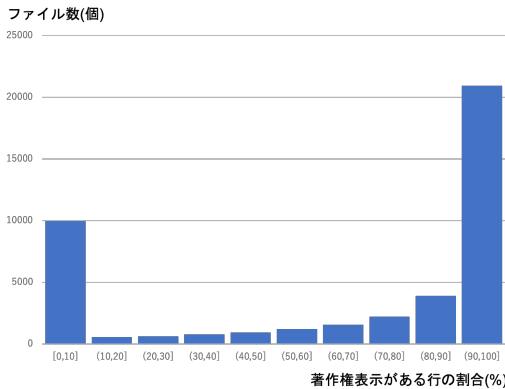


図 2 linux の著作権表示の割合別ファイル数

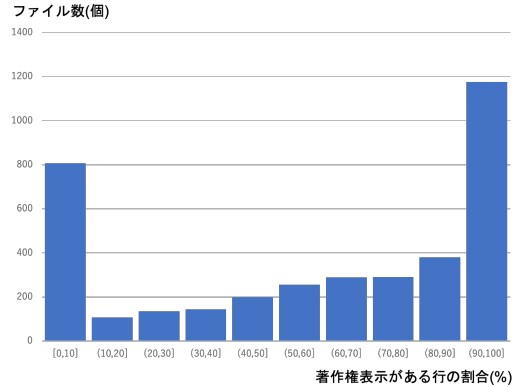


図 3 FFmpeg の著作権表示の割合別ファイル数

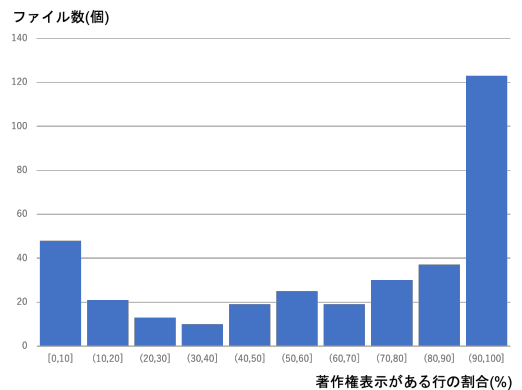


図 4 libuv の著作権表示の割合別ファイル数

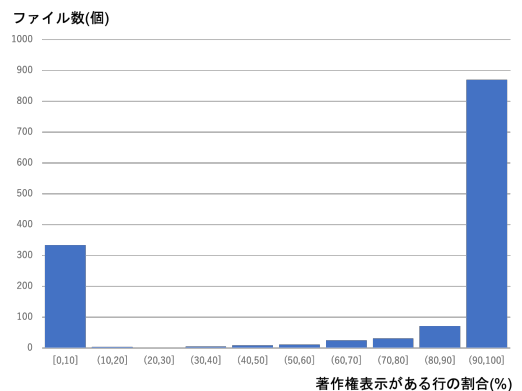


図 5 obs-studio の著作権表示の割合別ファイル数

- Step1: 対象ファイルのソースコードを確認し著作権表示を抽出する。
- Step2: git blame を用いて著作権表示の作成者のメールアドレスを取得する。
- Step3: git blame を用いて各行の作成者のメールアドレスを調べ、各行に対して著作権表示の有無を判定する。

4. 調査結果

4.1 OSS 全体の著作権表示の有無

各 OSS に対して全ての対象ファイルの全ての行の著作権表示の有無を調べた。その結果を表 2 にまとめる。

4.2 著作権表示の割合別ファイル数

各 OSS のファイルを著作権表示がある行数の割合で区別して、そのファイル数を調べた。調査結果を図 2 から図 6 に示す。図 2 から図 6 の各柱はファイルを著作権表示がある行数の割合で分類したものであり、ファイル中の X% の行に著作権表示があるファイルが Y 個あることを示している。

4.3 作成した行の割合と著作権表示の有無の割合

各 OSS の各ファイルに対して、そのファイルの全ての行に対するある組織または個人が作成した行の割

合と著作権表示を記述したかどうかの割合との関係調べた。調査結果を図 7 から図 11 に示す。図 7 から図 11 はファイル中の X% の行を作成した人または組織が Y % の割合で著作権表示を記述していることを示

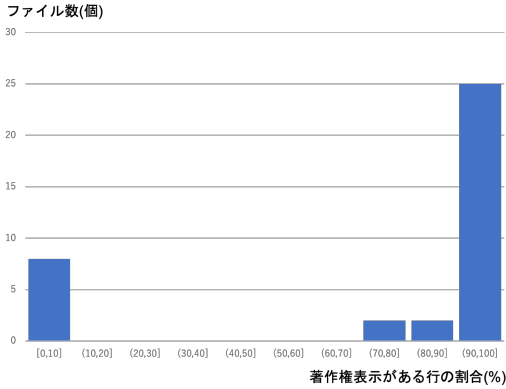


図 6 redis の著作権表示の割合別ファイル数

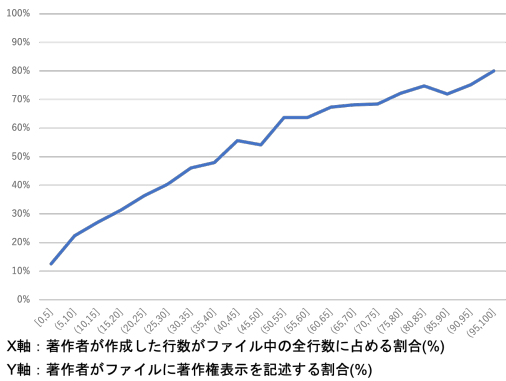


図 7 linux の作成した行の割合と著作権表示の有無の割合の関係

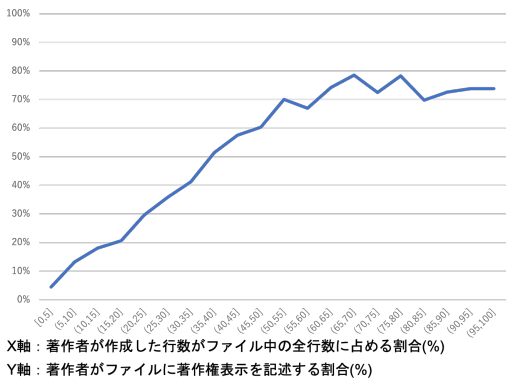


図 8 FFmpeg の作成した行の割合と著作権表示の有無の割合の関係

している。図 11 では redis の調査対象となったファイル数が少なかったため、存在しなかったデータの著作権表示を記述する割合は 0 となっている。

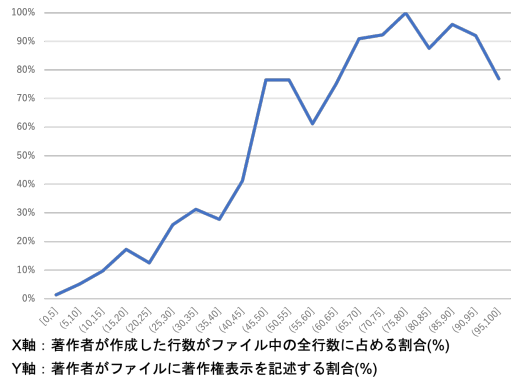


図 9 libuv の作成した行の割合と著作権表示の有無の割合の関係

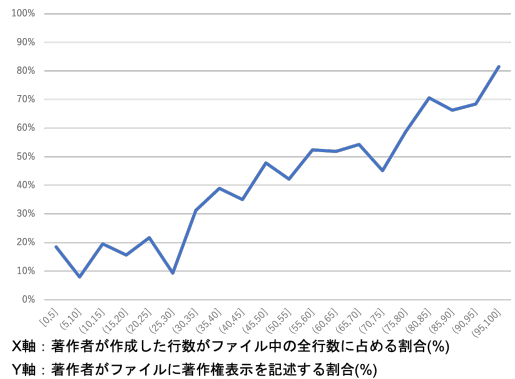


図 10 obs-studio の作成した行の割合と著作権表示の有無の割合の関係

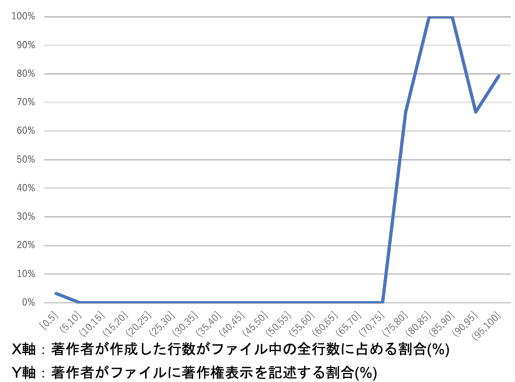


図 11 redis の作成した行の割合と著作権表示の有無の割合の関係

5. 考 察

表2の結果より、著作権表示がある行が半数程度またはそれ以上を占める一方、全ての行に著作権表示があるOSSはなかった。これは多くの人がOSSの開発に加わっている中で著作権表示に関するOSS全体での明確な合意がないことによるものだと考えられる。

また、全体の著作権表示がある行の割合はやや異なっていたものの、図2から図6の結果からは著作権表示の割合別ファイル数の分布傾向はほとんど変わらないように見えた。どのOSSにおいても、ほとんどの行に著作権表示があるファイルとほとんどの行に著作権表示がないファイルが最も多く、その中間のファイルはかなり数が少なかった。これは、ファイル作成時点では多くの場合、作成者が1人のため全ての行に著作権がある、もしくは全ての行に著作権表示がないのどちらかであるためだと考えられる。さらに、ほとんどの行に著作権表示があるファイルの方がほとんどの行に著作権表示がないファイルよりも多かったことから、ファイルの作成者は著作権表示を作成する人が多いことがわかる。

そして、図7から図11の結果もよく似ていた。どのOSSにおいてもファイルの全行数に占める作成した行数の割合が高いほど著作権表示を記述している割合が高いことがわかる。この結果から作成した行数が少ない人は著作権表示を記述しない可能性が高いことがわかる。また、折れ線グラフの変化の度合いが異なるのはファイル数によるものと考えられ、ファイル数の多いOSSになるほどなだらかに変化している。

6. 妥当性への脅威

本研究では作成者をメールアドレスの完全一致で分類しているため、作成者が一部でも異なる複数のメールアドレスを使用していた場合は同一人物または組織とはみなせず、著作権表示の有無が正しく判定できない場合がある。また、組織に所属しているかどうかをメールアドレスのドメインを用いて行なっているため、組織の職務として開発を行っているながらフリーのメールアドレスのドメインを使っている場合や、個人的に開発を行っているながら組織から支給されたメールアドレスのドメインを使っている場合には著作権表示の有無が正しく判定できない可能性がある。その上、本研究では使用したフリーメールアドレスに含まれていないフリーメールのドメインは正しく判定できない。さら

に、git blameは各行のAuthorのメールアドレスを表示するがAuthorとはコミットを追加した人を指すため実際の作成者とは異なる可能性がある。

7. ま と め

本研究ではいくつかのOSSに対して行単位で著作権表示の有無を定義し、その結果を調べた。まず、各OSSに対して全ての対象ファイルの全ての行の著作権表示の有無を調べた。その割合はOSSごとに大きく異なっていたものの、著作権表示がある行が半数程度またはそれ以上を占めていた。しかし、全ての行に著作権表示があるOSSはなかった。次に各OSSのファイルを著作権表示がある行数の割合で区別して、そのファイル数を調べた。その結果はどのOSSでもよく似ており、ほとんどの行に著作権表示があるファイルが最も多く、次に著作権表示がほとんどの行に存在しないファイルが多かった。さらに、各OSSの各ファイルの各組織または個人に対して、そのファイルの全ての行に対する作成した行の割合と著作権表示を記述したかどうかの割合の関係を調べた。その結果もどのOSSでもよく似ており、ファイルの全行数に占める作成した行数の割合が高いほど著作権表示を記述している割合が高いことがわかった。

今後の課題としては著作権表示がないソースコードが発生する原因を探るために開発者に著作権についてどういう考えを持っているかをアンケートすることやコミットを受けた際に著作権の扱いについて通知するツールの開発などが考えられる。

文 献

- [1] Open Source Initiative, The Open Source Definition,2007
- [2] Stewart, K. and Khan, S. and German, D.M., 2020 Linux Kernel History Report,2020
- [3] 著作権法（昭和四十五年法律第四十八号）17条2項・51条1項,2021
- [4] 著作権法（昭和四十五年法律第四十八号）21条~28条,2021
- [5] World International Property Organization, CRNR/DC/94 WIPO Copyright Title,1996
- [6] The World Trade Organization, TRIPS 協定,2017
- [7] 著作権法（昭和四十五年法律第四十八号）15条,2021
- [8] H. Meeker, Patrick mchardy and copyright profiteering,2017
- [9] D. M. German,B. Adams,K. Stewart, Cregit: Token-Level Blame Information in Git Version Control Repositories,2019
- [10] S. Qiu,D. M. German,K. Inoue, An Exploratory Study of Copyright Inconsistency in the Linux Kernel,2021
- [11] Chacon, S. and Straub, B., Pro Git,2020
- [12] GitHub, Inc., Where the world builds software,2021