

研究発表

Ninkaの分類困難なUnknownを分類 するための提案手法

2016 8/30発表

松浦 遊也

OSSの再利用

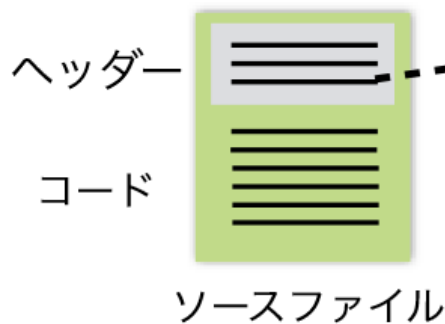
- 開発コスト削減のため、オープンソースソフトウェア（OSS）を再利用した派生開発が行われている



- OSSはソースファイル毎に決められている
- OSSライセンスを遵守することでOSSを再利用できる

ライセンス記述

- ライセンス記述はソースファイルの上部にコメントとして記述されている



```
#####  
# Redistribution and use in source and binary forms, with or without  
# modification, are permitted provided that the following conditions are met:  
#  
# 1. Redistributions of source code must retain the above copyright notice,  
#   this list of conditions and the following disclaimer.  
# 2. Redistributions in binary form must reproduce the above copyright  
#   notice, this list of conditions and the following disclaimer  
#   in the documentation and/or other materials provided  
#   with the distribution.  
#  
#####
```

- ライセンスを守るには、利用するソースファイル全てのライセンスを確認する必要がある

OSSのソースファイル数



Linux

ソースファイル数: 22310



ソースファイル数: 1452

- OSSのソースファイルは数が多く、全てのライセンスを手作業で確認するのは、労力がかかる

ライセンス特定ツールNinka[1]

- ライセンス記述を正規表現を用いてライセンス判定している



- 判定基準の記述はSPDXのリストに記述されているライセンス記述を用いている

[1]Ninka: Daniel M German, Yuki Manabe, and Katsuro Inoue, "A Sentence-Matching Method for Automatic License Identification of Source Code Files" In Proc. of ASE'10, pp. 437-446

SPDX

Giftware License	Giftware
GL2PS License	GL2PS
Glulxe License	Glulxe
GNU Affero General Public License v3.0	AGPL-3.0
GNU Free Documentation License v1.1	GFDL-1.1
GNU Free Documentation License v1.2	GFDL-1.2
GNU Free Documentation License v1.3	GFDL-1.3
GNU General Public License v1.0 only	GPL-1.0
GNU General Public License v2.0 only	GPL-2.0
GNU General Public License v3.0 only	GPL-3.0
GNU Lesser General Public License v2.1 only	LGPL-2.1
GNU Lesser General Public License v3.0 only	LGPL-3.0
GNU Library General Public License v2 only	LGPL-2.0
gnuplot License	gnuplot
gSOAP Public License v1.3b	gSOAP-1.3b

- 企業や組織がライセンスやコンポーネント情報を共有できるように基準を作成したもの

SPDX

GNU General Public License v2.0 only

Full name

GNU General Public License v2.0 only

Short identifier

GPL-2.0

Text

GNU GENERAL PUBLIC LICENSE

Version 2, June 1991

Copyright (C) 1989, 1991 Free Software Foundation, Inc.

51 Franklin Street, Fifth Floor, Boston, MA 02110-1301, USA

Standard License Header

Copyright (C) **yyyy name of author**

This program is free software; you can redistribute it and/or modify

This program is distributed in the hope that it will be useful, but WITHOUT
WARRANTY; see the GNU General Public License for more details.



基準となる
ライセンス記述

Ninkaの課題

- Unknownという未知のライセンスが出てしまう
- Unknownのライセンス内容を理解するためには、現状Unknownファイルを目視するしかない
- Unknownの結果が多いほど負担が大きくなる

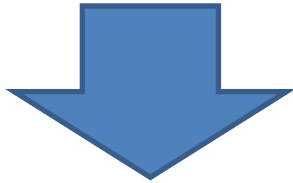
Unknownとなる原因

1. SPDXに存在するが対応できていない
2. 表記揺れに対応できていない



Ninkaが対応する可能性が高い

3. SPDXに登録される可能性が低いライセンス



種類は多いがそれぞれの数が少ないライセンス

後になっても対応される可能性が低い

研究目的

SPDXに登録される可能性が低い
ライセンスを分類できるようにすること

提案

SPDXに登録されていないUnknownは
ライセンスの制約毎に分類すればいいのでは？

予備調査

SPDXに登録されないライセンスには
どのような制約があるのか調査する

調査対象

- Free-bsd (Unknownライセンス記述1821)
 - SPDXにのっていないライセンス記述198
- Linux (Unknownライセンス記述3561)
 - SPDXにのっていないライセンス記述29

分類方法

- Can
 - 制限がなく、行ってもよいし、行わなくてもよい制約内容
- Cannot
 - 禁止されている制約内容
- Must
 - 必ず行わなければならない制約内容

結果1

- Can
 - 配布、使用
 - 著作権使用料無料使用
 - 商用使用
 - 修正
 - ソースフォームのみ配布
 - バイナリフォームのみ配布

結果2

- Cannot
 - 広告目的で(開発者名)を使用しない
 - サインなどに(開発者名)を使用しない
 - 許可なく(開発者名)を使用しない
 - ヘッダファイルを削除、変更しない
 - (開発した会社名)のマークを派生開発で使
用しない

結果3

- Must
 - ライセンス記述をそのままにする
 - 著作権をそのままにする
 - 免責事項をそのままにする
 - 「この製品は(開発した会社名)の製品を含む」という記載を入れる
 - バイナリでの配布の場合サポートドキュメントにライセンス記述を記載すること
 - このコメント文全てをそのままにする